Statistical Learning in Logistics and Manufacturing Systems

A Thesis
Presented to
The Academic Faculty

by

Ni Wang

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

School of Industrial and Systems Engineering
Georgia Institute of Technology
August 2006

Statistical Learning in Logistics and Manufacturing Systems

Approved by:

Professor Paul Kvam (Advisor)
School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Jye-Chyi Lu (Advisor)
School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor C. F. Jeff Wu
School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor John Vande Vate
School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Haizheng Li
School of Economics
Georgia Institute of Technology

Professor Ming Yuan
School of Industrial and Systems Engineering
Georgia Institute of Technology

Date Approved: 2 May 2006

To my wife, Jin Shi

To my parents

ACKNOWLEDGEMENTS

I would like to thank my two advisors: Dr. Paul Kvam and Dr. Jye-Chyi Lu. Their devoted guidance and support are the most valuable gifts I had during the journey of Ph.D. study. I would like to express my sincere gratitude to my advisor, Dr. Paul Kvam, for his prudent guidance and patient encouragement throughout the entire course of my research, and for his care and trust like a family member. He showed me the scholarly attitude toward research. I would like to acknowledge my advisor, Dr. Jye-Chyi Lu, for his contributions to my academic maturity. His great vision and leadership inspire and enable me to produce original research results. Beyond the technical expertise, I learned much more from him - the passion towards research, and the dedication to education.

I am very grateful to Dr. John Vande Vate for opening my eyes to the business world, and his support and recognition. I would like to thank Dr. C. F. Jeff Wu for his invaluable insights of my research work, his encouragement and recommendation. I also would like to express my grateful appreciation to Dr. Haizheng Li and Dr. Ming Yuan for their kindness to serve on my committee, and their valuable questions and comments.

There are so many friends and colleagues who have helped me during the past several years. I will also remember their friendship and support. And I would like to acknowledge the support from NSF grants DMI-0400071, DMS-0426056.

Finally, I would like to thank my wife, Jin Shi. She brings meaning to my life and makes this journey enjoyable to me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

This thesis focuses on the developing of statistical methodology in reliability and quality engineering, and to assist the decision-makings at enterprise level, process level, and product level.

In Chapter II, we propose a multi-level statistical modeling strategy to characterize data from large-size spatial logistics systems. The model can support business decisions at different levels, e.g., distribution center (DC) allocation at the strategic level covering large regions, exploring regional demand functions at the tactic level covering mid-size regions and demand forecasting at new store locations at the local level. The information available from higher hierarchies is incorporated into the multi-level model as constraint functions for lower hierarchies. A few analytical properties of the proposed methods are provided. Real-life examples from an automobile manufacture's logistics system demonstrate the importance of the research and illustrate the potentials of the proposed models. The key contributions include proposing the top-down multi-level spatial models which improve the estimation accuracy at lower levels; applying the spatial smoothing techniques to solve facility location problems in logistics.

In Chapter III, we propose methods for modeling system service reliability in a supply chain, which may be disrupted by uncertain contingent events. Given the large number of store locations and multiple combinations of routing schemes, this chapter applies an approximation technique for developing first-cut reliability analysis models. The approximation relies on multi-level spatial models to characterize patterns of store locations and demands. These models support several types of reliability evaluation of the logistics system under different probability scenarios and contingency situations. Examples with data taken from a large-scale logistics system of an automobile company illustrate the importance of supply-chain system reliability. The key contributions in this chapter are to bring statistical spatial modeling techniques to approximate store location and demand data, and to

build system reliability models entertaining various scenarios of DC location designs and DC capacity constraints.

Chapter IV investigates the power law process, which has proved to be a useful tool in characterizing the failure process of repairable systems. This chapter presents a procedure for detecting and estimating a mixture of conforming and nonconforming systems. The test of a mixture, based on a simple likelihood ratio, is illustrated with truncated failure data for copy machines. Bootstrap methods are used to gauge the estimation uncertainty, and optimal decisions for system replacement are determined based on the observed likelihood. The methods are applied in the analysis of copy machine failure data. The key contributions in this chapter are to investigate the property of parameter estimation in mixture repair processes, and to propose an effective way to screen out nonconforming products.

The key contributions in Chapter V are to propose a new method to analyze heavily censored accelerated life testing data, and to study the asymptotic properties. This approach flexibly and rigorously incorporates distribution assumptions and regression structures into estimating equations in a nonparametric estimation framework. The examples from industry of using available data to explore the regression functional relationship and distribution assumptions are provided. Derivations of asymptotic properties of the proposed method provide an opportunity to compare its estimation quality to commonly used parametric MLE methods in the situation of misspecified regression models. These examples and asymptotic studies show a significant potential of the proposed methodology.

CHAPTER I

INTRODUCTION

Statistical methods have been successfully applied in the areas of product design, process development, and other quality improvement activities. However, quality improvement should not be studied at the product level only. The increasing complexity and importance of logistics and manufacturing systems motivate us to develop new statistical methods and quality improvement techniques.

In Chapter II, we consider the spatial statistical methodology with applications at enterprise level business decisions. Significant uncertainty issues limit the application of traditional mixed-integer programming methods in solving complex large-scale problems of logistics management. Using spatial smoothing and interpolation methods, the uncertainty of customer locations and demands can be characterized. The spatial smoothing method can be integrated with continuum approximation approach to find design guidelines and solutions for large-scale facility location problems when the exact customer locations and demands are not available. Customer locations and demand data are important business drivers in planning logistics network. At different stages of business progress, the data available have different levels of granularity to support business decisions. It is important to combine the information from different sources to guide regional and local level decision making.

Chapter III focuses on the study of service reliability of logistics systems. The service reliability of such complex logistics systems depends on the strategy of how stores are supported by a distribution center (DC), and capability of the DC in replenishing materials to meet demands. However, there are cases where uncertainties can disrupt logistics operations. For example, the labor dispute at Long Beach in 2002 shut down the port for 10 days, with an estimated cost in the range of $2 billion - $20 billion. Despite the call for designing, and operating supply chains that are resilient to disruptions, there are few

analytical models for supply chain reliability. It is important to refine the framework and evaluation methods of supply chain system reliability, such that we can design and operate supply chains that are resilient to possible disruptions.

At the process level, Chapter IV discusses the repair and warranty management, which are also important in sustaining business growth, enhancing brand reputation, and building customer loyalty. Through investigating the repair history generated from the warranty process, the results help companies to characterize the product performance through statistical learning, and to make more informed decisions for future plans. For a repairable system, it is crucial to understand how the system's failure rate changes over usage time so that it can be taken off test before repairs become too frequent and too costly to the system operation. In many industries, the challenge is in finding which systems are susceptible to frequent failures and which ones are not. If a sound strategy can be provided to detect weak products as soon as possible, the warranty costs and operations costs can be greatly reduced.

Chapter V proposes and investigates a new quality improvement method at the product level. High-reliability product design and development begins with advanced product quality planning. Estimating the failure-time distribution or long-term performance of components product is difficult when few units will fail in experiments. Accelerated life testing (ALT) is commonly applied by stressing specimens at harsher conditions than in normal-use, thereby hastening failure time in tests with short duration. Typical parametric approaches assume that failure time distributions under various stress levels belong to the same parametric family and there is a (transformed) linear regression structure of the location parameters of these distributions. However, the parametric approach will lead to inefficient parameter estimates when the distribution function is misspecified. Nonparametric technique is a good alternative to analyze accelerated life-test data. We propose a nonparametric procedure by specifying regression relationships in estimating equations (EE), and then do the statistical inference on the regression parameters through the Empirical Likelihood (EL) method.

CHAPTER II

MULTI-LEVEL SPATIAL MODELING AND DECISION-MAKING WITH APPLICATION IN LOGISTICS SYSTEMS

2.1 Introduction

Supply-chain's logistics systems move and aggregate products from manufacturers at various sources (e.g., China) to global and local distribution centers (DCs) in several states and then to retailing stores. According to a recent report by Delaney (2004), the total logistics cost in the U.S. amounts to $936 billion in 2003, which is approximately 8.5% of the U.S Gross Domestic Product for that year. For improving supply-chain competitiveness, it is essential to utilize data representing customer shopping behaviors, product routing details, DCs' inventory situations, and so on to support decisions such as where to place DCs and stores so as to meet customer demands. Figure 1 presents an automobile manufacturer's domestic logistics network with a small number of DCs and numerous stores.

Customer locations and demand data are important business drivers in planning logistics network. At different stages of business progress, the data available have different level of granularity to support business decisions. Some of the decisions are more strategic, covering large geographical areas. Others are more tactical or operational focusing on local districts. For example, a retail chain company only have estimated numbers of potential stores serving its customers in respective market, without planning the exact locations of each individual store. Section 2 further motivates this research with examples of different levels of decisions including: DC locations analysis, demand driver analysis, and store demand predictions. Although the data in logistics management is rich, statistical techniques have not been thoroughly investigated to support various decisions. The goal of this article is to explore hierarchical data modeling techniques to support multi-level decisions in complex logistics

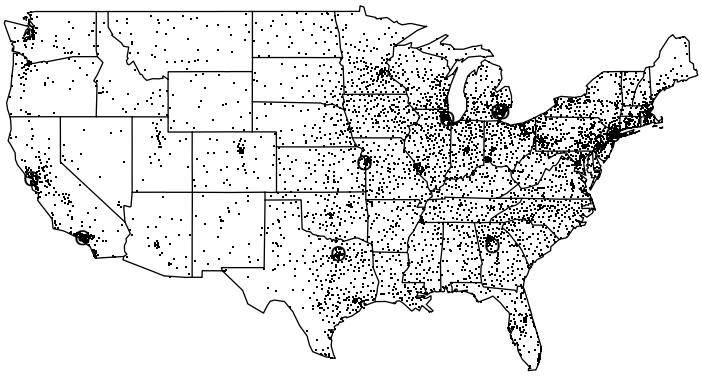

Figure 1: An Illustration of A Company's Domestic Logistics Network

systems.

The multi-level spatial models in this chapter focus on using aggregated data or executive decisions at higher levels to improve the prediction and inference at lower levels. The "level" has two-dimensional meaning: one dimension is "geographical" level such as national, regional and local; another dimension is time sensitive, referring to strategic, tactical and operational decisions. Synthesis of spatial data collected at different spatial scales is a familiar modeling problem, and has been extensively studied in statistical literature (Gotway and Young, 2002), this subject is known by different terms across various fields: multiscale and multi-resolution modeling, areal interpolation, modifiable areal unit problem, change of support problem, and so on. Data aggregated at higher levels will exhibit less sampling variation, but aggregated over smaller regions will offer a more detailed view of geographical patterns. Many approaches in the literature specify statistical models at fine spatial scales (e.g, a spatial point process) and study the results when data are grouped into higher scales, such as in block-kriging (Cressie, 1993). Typically, a specific multi-scale spatial model is proposed according to the formation of the multi-level hierarchy. For example, multiscale spatial tree models are used to describe spatial processes operating at multiple resolutions (see the review in Gotway and Young, 2002).

In some situations it is not possible to structure a single spatial model valid on all scales for different sources of data. In this article, we assume the data have been sampled, aggregated and processed before being made available for analysis. For example, for planning

DCs' serving areas, the global-models for supporting such strategic decisions can be built on "projected demand" data provided by regional managers. Wikle and Berliner (2005) used a Bayesian approach for solving this multi-scale and multi-source data modeling problem. They considered a conditional model by conditioning the spatially correlated process on an areal average of the process given from a higher hierarchy. In this way, the analyst avoids the challenging problem of constructing a model valid for all scales. This article presents an approach by using estimating equations as linkage to incorporate information from higher hierarchies to improve inference at lower levels. This approach is flexible (with less assumption like the conditional distribution) and effective in logistics planning. For example, local stores managers can make better store demand predictions by utilizing aggregated historic demand data available from the regional DC serving the stores.

Section 2 provides several motivating examples and describes the data used for illustration. Section 3 develops modeling techniques to combine data from store locations, store counts, population sizes and store sales for supporting these multi-level decisions. A new hierarchical modeling method using estimating equations to link key characteristics in models from different levels is proposed. See Section 4 for illustration of the examples from Section 2. Section 5 offers concluding remarks and further research tasks.

2.2 Examples of Data and Decisions at Different Scales

In solving large-scale supply-chain system resource-allocation problems, "continuum approximation" (CA) methods (Daganzo (1996), page 52) use smooth functions to describe store demand data or store locations at various locations and regions.

Denote $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{S}\}$ as a spatial point process for store demand and location. The set $\mathcal{S}$ is a collection of random events, and each realization is called a spatial point pattern. That is to say, we use a spatial point process $\mathcal{S}$ to model the store locations, and use random field $Y(\cdot)$ to represents the variable of interest such as store demand (e.g., Langevin *et. al.* 1996). For a given $\boldsymbol{s}_0$, assume that $Y(\boldsymbol{s}_0)$ is generated from a spatial random field $\{Y(\boldsymbol{s})\}$ independent of point process $\mathcal{S}$. Further assume that the point process $\mathcal{S}$ is a *nonhomogeneous Poisson Process* with intensity function $\lambda(\boldsymbol{s})$.

Most of the CA methods use a homogeneous Poisson process to model store locations within a region. The Poisson intensity function λ is usually obtained from the average number of stores. In practice the conventional CA method is not flexible enough to capture the different store location patterns in metropolitan and suburban areas. The following examples show that the proposed multi-level data modeling procedure enhances the CA method by improving its intensity estimation and also provides a better way to balance the model complexity and prediction accuracy for supporting decisions at different scales.

CA methods facilitate calculation of logistics costs for making strategic decisions about DC locations and serving areas. For example, Dasci and Verter (2001) studied the production-distribution system design issues. Erera and Daganzo (2003) considered stochastic load-constrained vehicle routing systems in which a fleet of homogeneous vehicles is dispatched from a central depot (e.g., a DC) to serve the demands of surrounding customers (e.g., retailing stores). Dandamudi and Lu (2003) used the CA method to derive an approximated logistics cost for supporting supply-chain contract decisions. Several difficult warehouse location-inventory-routing problems have been solved by using the CA method (see Daganzo and Erera (1999) for an example).

Example 1 (DC Location Analysis)

A Supply-chain's strategic decisions determine its logistics infrastructure for a longer-term period (more than 10 years). Facility-location allocation problems (Dasci and Verter, 2001) include deciding the optimal number, locations and serving areas of DCs in a region. The objective is to minimize the total costs as the summation of the fixed costs of operating DCs and the variable costs (e.g., transportation cost) for meeting store demands in the studied time window. To show the elements involved in these decisions, Geoffrion (1976) showed the optimal size of the service region is

$$|A|^* = \left(\frac{2F_C}{cK\lambda\mu_Y}\right)^{2/3}, \tag{1}$$

where F_C (\$/per year) is the (average) fixed cost of opening a DC, c is the (average) freight rate for shipments, λ is the intensity function of the homogeneous spatial Poisson process, μ_Y is the expected demand value for each store and K is a constant that depends on the

distance metric and the shape of the service region; this is explained in Langevelin, Mbaraga and Campbell (1996). Other elements such as truck travel speed, DC inventory-capacity, and so on can also be included in the model (1) for deciding the size of a DC's service region. Then, with the given total service region one can determine the number of DCs and allocate which DC is to serve which stores.

Table 2 shows the store counts of the 48 states in the continental United States. In Section 3.1, we extend the model using a nonhomogeneous Poisson process and develop a procedure to estimate the location-dependent intensity function. Figure 3 shows the smoothed intensity $\lambda(\cdot)$. In Section 4.1, we show how the extension of the model in (1) reflects the location-dependence in store demands. See Figure 4 for an example of the allocated DCs and their serving regions.

Example 2 (Demand Regression Analysis)

Tactical decisions in a supply chain include market analysis, distribution and transportation planning, production planning, and materials requirement planning. Unlike operational level decisions that are made during day to day operation, tactical decisions are updated several times every year, so refined models need to include more detailed information for capturing regional characteristics.

Most retailer companies want to know the sales performance and find ways to improve sales volume in regional markets. Market researchers studied the dependence of market performance and regional variables (Bronnenberg and Mahajan, 2001, Bronnenberg and Sismeiro, 2002). In the hierarchical sales network of the automobile manufacturer, suppose that regional sales executives want to study the relationship of store demand $Y(\cdot)$ and regional economical factors $x(\cdot)$ within a region B (e.g., the state of Texas). Due to sampling costs and other restrictions, several subregions $C_k \subset B$ are sampled to investigate the relationship of average store demand $Y(C_k$ and explanatory variables $x(C_k)$, where $k = 1, \ldots, n_0$. In this example, C_k's represent sampled counties in the state of Texas, and variable $x(C_k)$ $(= x_k)$ is available in aggregated form, such as county population, median household income, median house value, and so on. Meanwhile, from historical demand data

available at regional headquarter, the average store demand within region B is calculated to be $Y(B)$. As shown in Section 3.2, the upper level information $Y(B)$ can be used to improve the inference of the regional regression parameter in the regression analysis $Y(C_k) \sim x_k$.

Example 3 (Store Demand Prediction)

In local region, store managers are often concerned about the prediction of store demands at several possible candidate locations. From the regional analysis, the store demand at $\boldsymbol{s} \in C_k$ can be estimated by $Y_{opt}(\boldsymbol{s}) = \mathrm{E}Y(C_k)$. Moreover, to make more accurate small-scale predictions, it is better to consider the sales competitions from the stores nearby. Thus, a spatial model reflecting dependence between demands at nearby stores for projected future demands is needed. Denote $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_L\}$ as the locations for existing stores in this local district, and the store demands are given as $Y(\boldsymbol{s}_l)$, $l = 1, \ldots, L$. To predict future store demand $Y(\boldsymbol{s}_0)$ at a new site $\boldsymbol{s}_0$, the spatial dependence of $Y(\boldsymbol{s}_0)$ and existing $Y(\boldsymbol{s}_l$ need to be considered, and it will have an impact on prediction results. Section 3.3 presents prediction models and Section 4.3 contains corresponding numerical examples.

The remainder of this section provides more details of the problems discussed above and introduces notation for the subsequent model development. The spatial process is supported on a finite collection of non-overlapping subsets of the whole region A. Within A there are disjoint sets B_j, $j = 1, \ldots, n_b$. For example, A could be the continental U.S. and B_j's could be the states in U.S. In the next level, subregions C_{jk} (counties), $k = 1, \ldots, n_j$, are nested within B_j. Moreover, it is possible that the regions are not structured in a designed format. For example, following Example 1, a global DC's service region A_i (see Figure 5), $i = 1, \ldots, n_a$, could cover several states (or parts of states). The regional DC, then, could cover several counties within a state. In both cases there is a hierarchy in the multi-level system.

In a higher hierarchy where S^* represents a state B_j or a service region A_i, the (demand) data collected are in the aggregated form $Y(S^*)$ instead of the set of detailed realizations $\{Y(\boldsymbol{s}_i), \boldsymbol{s}_i \in S^*\}$. That is, although there is some linkage between the detailed data in the finer-scale spatial process $Y(\boldsymbol{s})$ and the aggregated data $Y(S^*)$ at the higher hierarchy, the

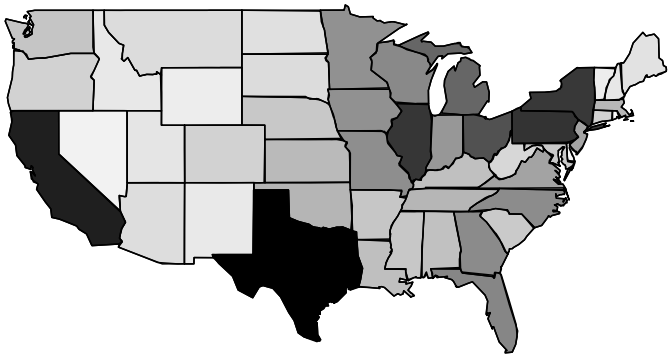

Figure 2: Counts $w(B_j)$ of Stores for each State (B_j) in USA (darker areas have more stores)

aggregated data is not the direct average (or sum) of all the detailed-scale data. In many real-life situations it is possible that not all the data are available from all store locations. For example, $Y(S^*)$ could be the average from a subset of sampled sites within region S, where the sampling scheme might not be known. It is also possible that $Y(S^*)$ represents the block-average in S, e.g., $Y(S) = |S|^{-1} \int_S Y(a)da$, based on a smooth function $Y(A)$ constructed from projected demands given by managers from the different regions, where $|S|$ is the area of $S \subset A$. See Section 4.1 for an example of a smooth function.

2.3 Multi-Level Model Developments

2.3.1 Global-Level Smoothing Models

In this section, we construct a global-level model for strategic planning decisions as introduced in Example 1. In CA analysis, one prerequisite is to understand the underlying store location pattern $\mathcal{S}$. By assuming a non-homogeneous Poisson process, it is sufficient to know the intensity function $\lambda(\boldsymbol{s})$. In the strategic planning stage, the data available are in the aggregate form: the number of stores to be opened in each market for the next decade. For simplicity reasons, we assume each state B_j is an independence market by itself ($j = 1, \ldots, n_b$). Figure 12 illustrates the aggregated total counts of stores for each state B_j in the whole region A, the continental United States, with the data contained in Table 2.

One simple way to obtain a spatial intensity function is to use kernel method (Cressie,

1993). However, in order to make sure that the estimated intensity function conforms with corporate planning. We impose the constraints $\int_{B_j} \lambda(\boldsymbol{s})d\boldsymbol{s} = \mathrm{E}W(B_j)$. Here, $\mathrm{E}W(B_j)$ is the expected number of stores in the subregion B_j projected through corporate expansion plan. Following the procedure proposed by Tobler (1979) and Dyn *et. al.* (1979). We consider the estimator $\hat{\lambda}(\boldsymbol{s})$ as the solution to the following constrained optimization problem,

$$\iint_B \left[\left(\frac{\partial^2 \lambda}{\partial u^2} \right)^2 + 2 \left(\frac{\partial^2 \lambda}{\partial v \partial u} \right)^2 + \left(\frac{\partial^2 \lambda}{\partial v^2} \right)^2 \right] dudv \tag{2}$$

subject to the constraints $\lambda(\boldsymbol{s}) \geqslant 0$ and

$$\int_{B_j} \lambda(\boldsymbol{s})d\boldsymbol{s} = \mathrm{E}W(B_j), \text{ for } j = 1, \ldots, n_b, \tag{3}$$

where $\boldsymbol{s} = (u, v)$ and (u, v) is the two-dimensional coordinate for the region A. The constraint in (29) preserving the total volume in each subregion B_j is called the *pycnophylactic property* (or mass preserving property) by Tobler. The following finite difference method can be applied to solve this constrained optimization problem, and the existence and uniqueness of the solution are proven by Tobler (1979).

First, a fine mesh of equally spaced points is superimposed over the entire region A. The value of the smooth function $\lambda(\boldsymbol{s})$ is approximated by the values calculated at these mesh points using procedures (28) and (29). We denote by r and c the row and column indices for the mesh lattice, respectively, and $\lambda_{r,c}$ for the value for $\lambda(\boldsymbol{s})$ when $\boldsymbol{s}$ is in the mesh block $b(r, c)$. The derivative of the smooth function in (28) is then approximated by

$$\frac{\partial^2 \lambda}{\partial u^2} \approx (\lambda_{r,c+1} - 2\lambda_{r,c} + \lambda_{r,c-1})/(\Delta u)^2$$

and the mass-preserving constraint is enforced by requiring that

$$\Delta u \Delta v \sum_{b(r,c) \in B_j} \lambda_{r,c} = \mathrm{E}W(B_j).$$

Thus, the smooth function of $\lambda(\cdot)$ is approximated by discrete values from the mesh grids. See Tobler (1979) for more details on implementing the procedure. The estimation quality from different choices of mesh sizes was not fully discussed; this issue deserves careful attention, but is outside the focus of this chapter.

Similarly, if the average demand $\mathrm{E}Y(B_j)$ is given, the intensity function $\hat{Y}(\boldsymbol{s})$ for $\boldsymbol{s} \in B_j$ can also be calculated under the constraints

$$\frac{1}{|B_j|}\int_{B_j} Y(\boldsymbol{s})d\boldsymbol{s} = \mathrm{E}Y(B_j), \text{ for } j = 1, \ldots, n_b. \tag{4}$$

Note that there is no distributional assumption for Y other than preserving its means across subregions. An application of the smooth intensity functions in a strategic decision of the number of global DCs and their serving areas is detailed in Section 4.1.

2.3.2 Regional-Level Hybrid Models of Likelihood and Regression

This section focuses on one region B_0 and its subregions C_{0k}, where $k = 1, \ldots, n_0$. Hereafter, the subscript 0 is omitted for brevity. Similarly, for notational convenience the average demand $Y(C_k)$ and store counts $W(C_k)$ are abbreviated by Y_k and W_k, respectively. Assume that W_k is a Poisson random variable with mean λ_k, and the maximum likelihood estimates (MLEs) for the mean parameters are $\hat{\lambda}_k = w_k$, where w_k is observed store counts at each county, $k = 1, \ldots, n$.

To study the effect of a regional economic covariate x_k on the demand Y_k, the (average) demands are sampled from m counties within a state. Denote $f(y_k; \beta, x_k)$ as the conditional probability density function (pdf) of y_k, where β is a vector regression parameters for covariates x_k. Note that covariate measures here can include county population, buying-power indicators, regional DC locations and their serving areas, personnel requirements for each regional DC, and so on.

Suppose $m < n$ subregions are sampled from the n available ones. Without loss of generality, denote the available data as $\{y_k, w_k\}, k = 1, \ldots, m$ and $w_{m+1}, \ldots, w_n$. Note that although the store counts at counties C_k, $k = m+1, \ldots, n$, are given, demands are not available due to the high cost of obtaining such data.

Conditional on regional covariate x_k, assume store counts W_k and county demands Y_k are independent. Thus, their joint likelihood is:

$$L(\boldsymbol{\lambda}, \beta) \propto \left(\prod_{k=1}^{n} \frac{\lambda_k^{w_k}\exp(-\lambda_k)}{w_k!}\right)\left\{\prod_{k=1}^{m} f(y_k; \beta, x_k)\right\}. \tag{5}$$

This hybrid model is proposed for combining information of store counts, (average) demand data and covariates supporting regional-level tactical decisions. It is natural to expect that the total store count and demand) in a specific region B obtained from regional level match with the information from higher level. More specifically, the mean for the total counts and expected value for average demand data in region B from a higher level are $\lambda_B = \mathrm{E}W(B)$ and $\mathrm{E}Y(B)$, respectively, as discussed in Section 3.1. On the other hand, from the lower level λ_B will be the sum of individual λ_k from subregions C_k with $k = 1, \ldots, n$. The expected demand can be obtained similarly. Thus, linkage across the two hierarchies can be summarized as:

$$\sum_{k=1}^{n} \lambda_k = \lambda_B, \quad \sum_{k=1}^{n} \lambda_k \cdot \mathrm{E}Y_k = \lambda_B \cdot \mathrm{E}Y(B). \tag{6}$$

The above two equations form natural constraints for estimating model parameters. Other constraints concerning higher distribution moments can also be entertained (but is not studied here).

The *proportional intensity* is defined to be $p_k = \lambda_k/\lambda_B$, where $0 \le p_k \le 1$ and $\sum_k p_k = 1$. In terms of $\mathbf{p}' = (p_1, ..., p_n)$, the hybrid likelihood (5) can be simplified to:

$$L(\mathbf{p}, \beta) \propto \left(\prod_{k=1}^{n} p_k^{w_k} \right) \left\{ \prod_{k=1}^{m} f(y_k; \beta, x_k) \right\}. \tag{7}$$

In the regional-level model, consider the expectation $\mathrm{E}Y_k$ as a function of covariates and their associated regression parameters. Then, $g(x_k, \beta) = \mathrm{E}Y_k - \mathrm{E}Y(B)$ for $k = 1, \ldots, n$ form the estimating equations that provide the covariate information. Thus, the constraints in (6) can be expressed as:

$$\sum_{k}^{n} p_k = 1, \quad \sum_{k=1}^{n} p_k g(x_k, \beta) = 0. \tag{8}$$

New estimators $(\tilde{\boldsymbol{\lambda}}, \tilde{\beta})$ can be obtained through maximizing the likelihood function in (7) under the constraints in (8). The likelihood L can be regarded as a hybrid of the nonparametric likelihood $\prod p_k^{w_k}$ and the parametric likelihood with pdf $f(y_k)$ using the auxiliary information as constraints. This constrained maximization setup follows the empirical likelihood approach (Qin and Lawless (1994), Owen (2001)) as a natural way to improve the statistical inference by incorporating auxiliary information through estimating equations.

Qin (2000) demonstrated that parametric and empirical likelihood can be combined to produce valid inferences for paired data. Note that in (7) the empirical probability p_k has a power term w_k, which acts as a weighting function to p_k. This is quite different from the above three references where the power terms are all equal to one.

Estimation of parameters in the hybrid likelihood (7) with constraints (8) can be accomplished through optimizing the following Lagrange objective function, where ν and τ are the Lagrange multipliers:

$$H = \sum_{k=1}^{n} w_k \log p_k + \sum_{k=1}^{m} \log f(y_k; x_k, \beta) + \nu(1 - \sum_{k=1}^{n} p_k) - n\tau \sum_{k=1}^{n} p_k g(x_k, \beta). \tag{9}$$

By taking derivatives with respect to p_k and setting them to zero, we have:

$$p_k = \frac{w_k}{\sum_k w_k + n\tau g(\boldsymbol{x}_k, \beta)},$$

where τ is determined in terms of β through the equation

$$0 = \sum_k p_k g(x_k, \beta) = \sum_k \frac{w_k g(x_k, \beta)}{\sum_k w_k + n\tau g(x_k, \beta)}. \tag{10}$$

Utilizing these p_k's and taking the logarithm of L in (7), the profile log-likelihood of β is

$$l(\beta) = \sum_{k=1}^{n} w_k \log \left(\frac{w_k}{\sum_k w_k + n\tau g(x_k, \beta)} \right) + \sum_{k=1}^{m} \log f(y_k; \beta, x_k). \tag{11}$$

The constrained MLEs of β can be obtained numerically from solving the following equations:

$$\partial l(\beta)/\beta = -\sum_{k=1}^{n} \frac{n\tau w_k}{\sum_k w_k + n\tau g(x_k, \beta)} \frac{\partial g(x_k, \beta)}{\partial \beta} + \sum_{k=1}^{m} \frac{\partial}{\partial \beta} \log f(y_k; x_k, \beta) = 0. \tag{12}$$

To keep the presentation brief, the large sample properties of the estimates are left to the Appendices. See Section 4.2 for numerical examples.

2.3.3 Local-Level Kriging Models

When the demand characteristics of individual stores are of interest (see Example 3), inference on the random field $Y(\boldsymbol{s})$ is needed at a more detailed level. Consider the locations within a local district $C \subset B$ as a collection of several neighboring counties, $C = \bigcup C_k$. Due to the data aggregation and sampling processes, the spatial dependence of "block averages" of demands is difficult to model; a full specification of a complete model of $\{Y(\cdot)\}$

at all scales is difficult and complex. To overcome this difficulty, Wikle and Berliner (2005) defined a continuous process $Y(\boldsymbol{s}) = Y(C_k) + \delta(\boldsymbol{s})$ for $\boldsymbol{s} \in C_k$, which has the following properties: conditional on $Y(C_k)$, $\delta(\boldsymbol{s})$ is a spatial process with mean 0, and covariance function $\mathrm{Cov}(\delta(\boldsymbol{u}), \delta(\boldsymbol{v}))$. Note that in this formulation the spatial dependence is only focused on the detailed level for store locations within C_k. The spatial dependence at the high hierarchical locations becomes a nuisance when using a conditional-mean $Y(C_k)$ treatment. Wikle and Berliner adopted a Bayesian approach to specify the conditional distributions on each hierarchy for model inference. Under their assumption, with $\boldsymbol{s} \in C_k$, $\mathrm{E}Y(\boldsymbol{s}) = \mathrm{E}Y(C_k)$ and $\mathrm{Cov}(Y(\boldsymbol{u}), Y(\boldsymbol{v})) = \mathrm{Cov}(\delta(\boldsymbol{u}), \delta(\boldsymbol{v})) + \mathrm{Cov}(Y(C_u), Y(C_v))$, for $\boldsymbol{u} \in C_u$ and $\boldsymbol{v} \in C_v$, where $\boldsymbol{u}$ and $\boldsymbol{v}$ are locations in C_k.

However even the conditional distribution is hard to specify in some cases, especially when the aggregating processes for data available from different hierarchies are not known. We consider the multi-level spatial models satisfying the following two conditions:

1. After a proper transformation of $Y(\cdot)$, the transformed store demand $Z(\cdot)$ satisfy $Z(\boldsymbol{s}) = \mathrm{E}Z(C_k) + \delta(\boldsymbol{s})$.

2. $\delta(\boldsymbol{s})$ is a Guassian process with mean 0, and covariance function $\mathrm{Cov}(\delta(\boldsymbol{u}), \delta(\boldsymbol{v})) = \boldsymbol{C}_\delta(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{C}_\delta(\boldsymbol{u} - \boldsymbol{v})$, where the last equality ensures the second-order stationary.

In other words, the data (in terms of block averages) collected at the upper hierarchies are used as references or constraints for the processes at the lower hierarchies in the multi-level systems. The variations within the region B are captured by the spatial (local-)dependence $\boldsymbol{C}_\delta(\boldsymbol{u}, \boldsymbol{v})$ for $\boldsymbol{u}, \boldsymbol{v} \in C$.

As seen in Example 3, a new store plan at location $\boldsymbol{s}_0 \in C \subset B$ depends on our ability to predict the store demand $Y(\boldsymbol{s})$. By combining the given block-average demands $Y(C_k)$, and the detailed demand observations $Y(\boldsymbol{s}_l)$ for stores at the neighborhood locations $\boldsymbol{s}_l \in C$, $l = 1, \ldots, L$, the spatial prediction of $Y(\boldsymbol{s}_0)$ can be obtained using a slight modification of the simple Kriging method (Cressie 1993, Chapter 3) explained below.

Let's focus on the transformed Z scale, and the Kriging problem can be summarized as follows: given $Z(\boldsymbol{s}_l), l = 1, \ldots, L$, one wants to predict $Z_0 = Z(\boldsymbol{s}_0)$ at a nearby location $\boldsymbol{s}_0$.

According to the assumption of our spatial model, $Z(\boldsymbol{s}_0) = \mathrm{E}Z(\boldsymbol{s}_0)+\delta(\boldsymbol{s}_0) = \mathrm{E}Z(C_0)+\delta(\boldsymbol{s}_0)$ for $\boldsymbol{s}_0 \in C_0$. Since $\mathrm{E}Z(C_0)$ can be calculated based on regional level regression model, we only need to focus on the prediction of the residual $\delta(\boldsymbol{s}_0)$. Given $\boldsymbol{Z} = \{Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_L)\}$, the residuals $\boldsymbol{\delta} = \{\delta(\boldsymbol{s}_1), \ldots, \delta(\boldsymbol{s}_L)\}$ can be obtained by removing the mean effects. Under squared-error loss, the best linear unbiased predictor for $\delta(\boldsymbol{s}_0)$ is $\mathrm{E}(\delta(\boldsymbol{s}_0)|\boldsymbol{\delta})$. Under the Gaussian Process assumption for $\delta(\cdot)$, the optimal predictor for $\delta(\boldsymbol{s}_0)$ is

$$\delta_{opt}(\boldsymbol{s}_0) = \boldsymbol{c}_\delta{}'\Sigma_\delta{}^{-1}(\boldsymbol{\delta} - \boldsymbol{\mu}) + \mu(\boldsymbol{s}_0), \tag{13}$$

where $\boldsymbol{c}_\delta = (\boldsymbol{C}_\delta(\boldsymbol{s}_0, \boldsymbol{s}_1), \ldots, \boldsymbol{C}_\delta(\boldsymbol{s}_0, \boldsymbol{s}_L))$, $\boldsymbol{\mu} = (\mathrm{E}\delta(\boldsymbol{s}_1), \ldots, \mathrm{E}\delta(\boldsymbol{s}_L))$, $\mu(\boldsymbol{s}_0) = \mathrm{E}\delta(\boldsymbol{s}_0)$ and Σ_δ is an $L \times L$ matrix whose (i,j)th element is $\boldsymbol{C}_\delta(\boldsymbol{s}_i, \boldsymbol{s}_j)$. When $\boldsymbol{\mu}$ and $\boldsymbol{C}_\delta(\boldsymbol{s}_i, \boldsymbol{s}_j)$ are unknown, efficient estimates are used in place of the corresponding unknown parameters in the optimal predictor to construct the spatial predictors.

The covariance matrix Σ_δ and the vector $\boldsymbol{c}_\delta$ can be estimated by exploring the dependence structure (e.g., semi-variogram) between observations at locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_L$ in the neighborhood of $\boldsymbol{s}_0$. Define the semi-variogram $\gamma_\delta(\boldsymbol{u}, \boldsymbol{v})$ for the (demand) observations at two nearby locations as:

$$2\gamma_\delta(\boldsymbol{u}, \boldsymbol{v}) = \mathrm{Var}(\delta(\boldsymbol{u}) - \delta(\boldsymbol{v})).$$

Assuming second-order stationarity for $\delta(\cdot)$, the variogram is then related to the covariance function through

$$2\gamma(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{C}_\delta(\boldsymbol{u}, \boldsymbol{u}) + \boldsymbol{C}_\delta(\boldsymbol{v}, \boldsymbol{v}) - 2\boldsymbol{C}_\delta(\boldsymbol{u}, \boldsymbol{v}). \tag{14}$$

Cressie (1993) discusses several parametric variograms and the associated estimation methods of model parameters. With the estimates of the semi-variogram, the estimates of Σ_δ and $\boldsymbol{c}_\delta$ are readily obtained, and the estimation procedure of $\delta_{opt}(\boldsymbol{s}_0)$ in (13) can be completed.

2.4 Application of Multi-Level Models

2.4.1 Facility-Allocation Using Large-Scale Models

In the global-level logistics planning problem as seen in Example 1, the decision variables are the optimal number (n_a) of national DCs, the DC locations, $\{\boldsymbol{s}_{A_i},\ i = 1, \ldots, n_a\}$, and respective service regions $\{A_i, i = 1, \ldots, n_a\}$. Assume that a single DC is allocated

to satisfy all the store demands in its service region. Although service regions may have irregular shapes different from the circles, hexagons, or squares suggested in the economics and transportation literature, this irregularity is shown to have little impact on the optimal solution (Geoffrion, 1976).

For simplicity, our illustration assumes that the DC has unlimited capacity to handle requested demands (Dasci and Verter 2001), and there is no operating and maintenance costs for DCs. Then, the total logistics cost $\mathcal{C}_T$ is comprised of the fixed costs for DCs and variable transportation costs $\mathcal{C}_v$ delivering products to meet demands. Denote $\mathcal{C}_f(\boldsymbol{s}_{A_i})$ (\$/per year) be the fixed "infrastructure" cost of opening a DC at $\boldsymbol{s}_{A_i} \in A$, and $\mathcal{C}_{tr}(\boldsymbol{s}_{A_i}, A_i)$ (\$/per year) be the total transportation cost for the DC location $\boldsymbol{s}_{A_i}$ and service region A_i. The total logistics cost is $\sum_i^{n_a}\{\mathcal{C}_f(\boldsymbol{s}_{A_i}) + \mathcal{C}_{tr}(\boldsymbol{s}_{A_i}, A_i)\}$. Suppose that the transportation costs are charged on a per item-mile basis, and denote $c(\boldsymbol{s})$ the freight rate for shipments originating from $\boldsymbol{s}$. Let $K(\boldsymbol{s})$ be the constant that depends on the distance metric and the shape of the service region, which is explained in Langevin *et al.* (1996). For approximately circular service regions with the facility at the center, Dasci and Verter (2001) showed that the total transportation costs can be expressed as

$$\sum_i \mathcal{C}_{tr}(\boldsymbol{s}_{A_i}, A_i) = \sum_i c(\boldsymbol{s}_i)K(\boldsymbol{s}_i)\lambda(\boldsymbol{s}_{A_i})\mu_Y(|A_i|)^{3/2}, \tag{15}$$

where $\lambda(\cdot)$ is the intensity function of the nonhomogeneous spatial Poisson process , $|A_i|$ is the area of the service region A_i, and μ_Y is the expected demand value for each store in A_i. The proposed DC location-allocation design problem is formulated as follows:

$$\begin{aligned} &minimize: \quad \mathcal{C}_T = \sum_i \{\mathcal{C}_f(\boldsymbol{s}_i) + \mathcal{C}_{tr}(\boldsymbol{s}_i, A_i)\} \\ &subject\ to \quad \bigcup_i A_i = A,\ \boldsymbol{s}_i \in A_i, \end{aligned} \tag{16}$$

where A_i's are disjoint. By following the work in Geoffrion (1976) and Dasci and Verter (2001), without showing the details, the optimal size of service regions can be derived as follows:

$$|A(\boldsymbol{s})|^* = \left(\frac{2\ \mathcal{C}_f(\boldsymbol{s})}{c(\boldsymbol{s})K(\boldsymbol{s})\lambda(\boldsymbol{s})\mu_Y}\right)^{2/3}. \tag{17}$$

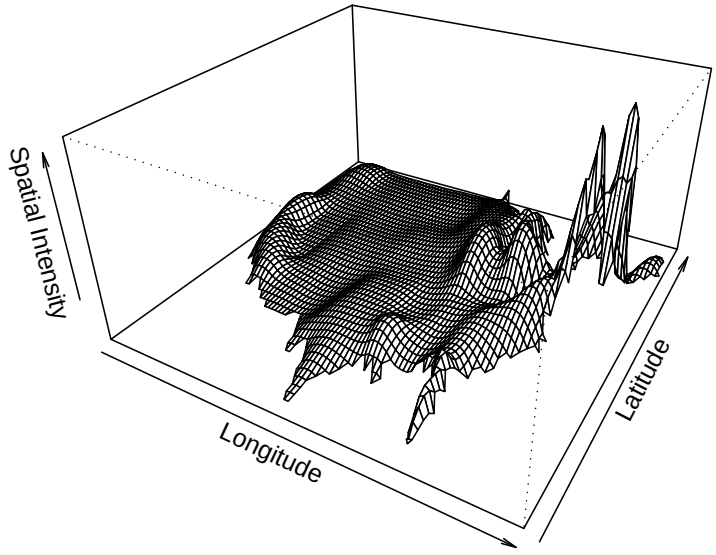

Figure 3: Contour of Intensity Function from the Global-Level Model

Note that this result is a simple extension of (1) with all the elements (except the expected demand in each store μ_Y) dependent on the location $\boldsymbol{s}$.

For the purpose of illustrating a supply-chain's logistics system-design study, we consider data and cost parameters contributed from a major U.S. automobile retailing company. For proprietary reasons, data and costs are adjusted to disguise the true costs and operational parameters. Store counts in Table 2 are provided in an aggregated form for the 48 contiguous states. Based on the count data $W(B_j)$, a smooth function $\lambda(\boldsymbol{s})$ is obtained using the constrained interpolation method discussed in Section 3.1, where a 100×50 grid is overlaid on the U.S. map. All the computations are implemented using R (R Development Core Team, 2004). Figure 2.4.1 presents the contour for the intensity function of store counts. Note that the New York and Boston metropolitan areas have the darkest intensities representing the most populated regions.

In applying the CA method to solve facility location-allocation problems (e.g., Ouyang and Daganzo, 2003) it is generally assumed that the intensity function is known and continuous. However, when applying the smoothing technique described in Section 3.1 to estimate the intensity function, the function is discrete subject to the grid size considered. That is,

the intensity values for nearby grid-blocks will be discontinuous with some "jumps" between blocks. The following extends the iterative algorithm in Ouyang and Daganzo (2003) by evaluating the optimal service-region size (17) at grid-based lattice locations to solve the DC location-allocation problem.

1. Starting Points: Assume that the stores are distributed uniformly over the whole region A. By using the averages of all cost parameters and the intensity function over all subregions, (1) provides an initial size $|A_i|^{(0)}$ of the service regions. The number $(n_a^{(0)})$ of DCs is then $|A|/|A_i|^{(0)}$. A is covered using $n_a^{(0)}$ circular disks with area $|A_i|^{(0)}$ such that these disks do not overlap. Locate the DC at the geographical centers $\boldsymbol{s}_{A_i}^{(0)}$ of $A_i^{(0)}$.

2. Locate $\boldsymbol{s}_{A_i}$ in the superimposed lattice such that $\boldsymbol{s}_{A_i} \in b(r_i, c_i)$. Apply the interpolated value $\hat{\lambda}(r_i, c_i)$ available from the superimposed lattice in (17) to calculate the optimal area of service regions $|A_i|$. Move the disks manually (or via an automatic algorithm) in a way that they do not overlap, and locate $\boldsymbol{s}'_{A_i}$ as the new geographical centers for the service region A_i. Note that every time the center moves, the value of the estimated intensity $\hat{\lambda}(\cdot)$ evaluated at this new center will change. The size $|A_i|'$ calculated at this new center location will also change.

3. Go to Step 2 to recalculate of the area of service regions $|A_i|'$ based on $\hat{\lambda}(\cdot)$ evaluated at the new block $b(r_i', c_i')$, where $\boldsymbol{s}'_{A_i} \in b(r_i', c_i')$.

4. Repeat steps 2-3 until the process reaches an acceptable solution that none of the service regions overlap and $\bigcup A_i^* \approx A$.

Returning to the auto retail company problem, suppose the fixed cost $\mathcal{C}_f(\boldsymbol{s})$ for establishing a new national DC is set as \$10,000 per day. Transportation costs are charged on a per-item-mile basis, with $c(\boldsymbol{s}) = c = \$0.004$ per item per mile. $K(\boldsymbol{s}) = K$ depends on the distance metric and the shape of the service region. Values of K for commonly used service-region shapes and metrics can be found in Dasci and Verter (2001). For the Euclidean distance and circular service region, $K = 0.3$. There are in total $M = \sum_{j=1}^{n_b} \mathrm{E}W(B_j) = 5216$

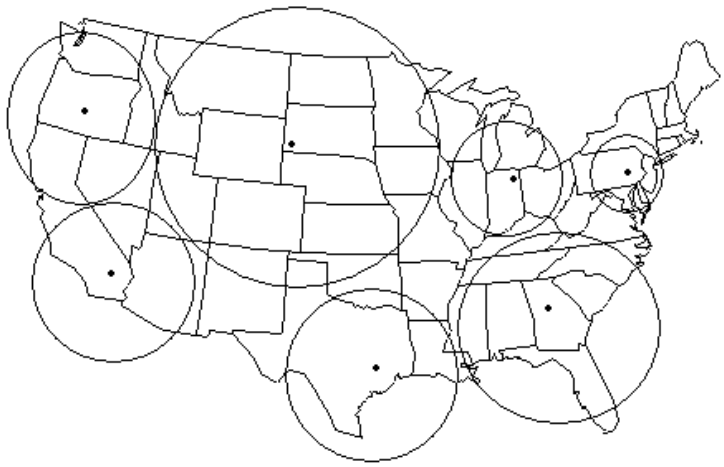

Figure 4: An Illustration of the DC Allocation Process

stores in the 48 states. To find the starting values, we initially assume the stores are distributed uniformly over the whole region A. The constant intensity for the spatial Poisson process is $\lambda(\boldsymbol{s}) = \lambda^{(0)} = M/|A|$, where $|A| = 3,537,438$ square miles for the continental U.S. and the average demand is given as $\mu_Y = 40$ items per day per store. The optimal area of service region is then $A_i^{(0)} \approx 445766$ square miles. Thus $A/A_i{}^* \approx 8$ non-overlapping DCs are initially allocated to serve the continental U.S.

Going through the algorithm presented above, Figure 4 shows the DC service regions with the DC located in the center of the region. Note that since the values of the intensity function $\lambda(\cdot)$ from grids in the Northeast region are about 8 times the intensity values for grids from the Southwest region, the corresponding radius of the service region serving Northeast region is approximately 1/2 of the radius of the service region serving Southwest region according to (17). The method illustrated here provides a way to evaluate different strategies in structuring a supply-chain's logistics systems.

Remarks:

1. For simplicity, the algorithm above assumes constant cost parameters and average demand. When the (smooth) functions of cost parameters and demands are given, procedures similar to $\hat{\lambda}(\cdot)$ evaluated at discrete blocks can be used to make the decision of service regions more realistic.

2. As in Ouyang and Daganzo (2003), the algorithm above uses a single evaluation of $\hat{\lambda}(\cdot)$ at the new center points $\boldsymbol{s}'_{A_i}$ for the calculation of the size $|A_i|'$. Further research is needed to evaluate its appropriateness and to explore ways for using the average from the center's neighborhood (although there is a problem in deciding neighborhood-size) for deciding the size $|A_i|'$.

3. The quality of estimating $\lambda(\cdot)$ (e.g., bias and variance) will affect the decision of DCs' service regions (and the number of DCs). When the size of a service region is too large than its necessity, this will create a "stock-out" problem (no product to sell in stores), which is a major cost challenge in retailing-chain operations. Further research is needed to understand ways to include estimation quality into the decision formulas such as (17).

2.4.2 Regional-Scale Demand Analysis

In this section, we consider the state of Texas, represented as B, as the primary study interest. The goal is to explore the relationship between the county-level average-demand $Y(C_k)$ and covariates such as population size, buying power, etc. evaluated at the county level. The store counts $w(C_k)$ are available for all counties C_k ($k = 1, \ldots, 255 = n$) in Texas. A random sample of $m = 50$ (average) county-level demands ($Y(C_k)$, $k = 1, \ldots, m$) is collected within Texas for regression analysis. Logarithms transformation is applied (i.e., $Z(C_k) = \log(1 + Y(C_k))$) before conducting a regression analysis to normalize the original data. After fitting several covariate variables including population size and buying power (represented by median household income and median house value) and evaulating their statistical significance, we choose $x_k = \log(\text{county population size})$ to be the regional covariate for county C_k in region B for this illustration example.

Through exploratory analysis, assume a linear model on the covariate effect, $Z(C_k) = \beta_0 + \beta_1 x_k + \epsilon_k$, where $\epsilon_k \sim N(0, \sigma_z^2)$, $k = 1, \ldots, m$. Refer to constraint functions in (8) constructed from a higher hierarchy and the bias-adjustment (19) in the log-transformation

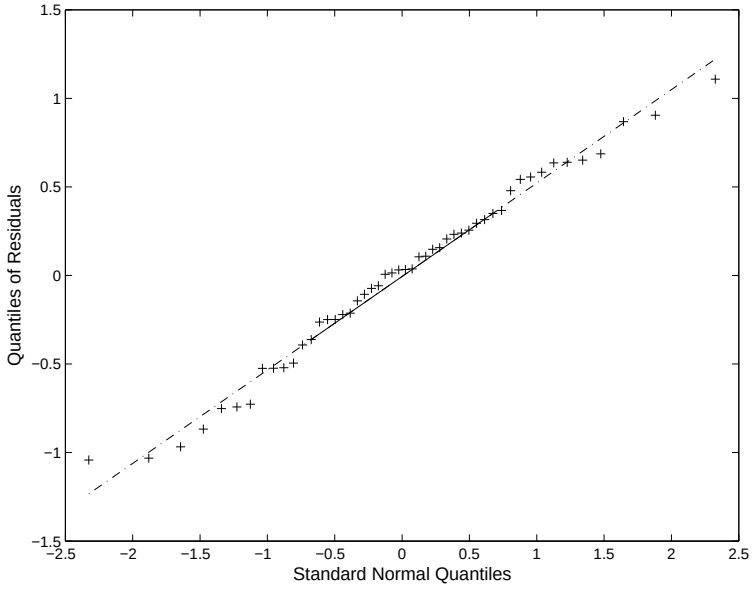

Figure 5: QQ Plot of Regression Residuals

for normality. The constraint functions on the average-demand then become

$$\begin{aligned}\sum_{k=1}^{n} p_k g(x_k, \beta) &= \sum_{k=1}^{n} p_k \left\{ \mathrm{E}Y(C_k) - \mathrm{E}Y(B) \right\} \\ &= \sum_{k=1}^{n} p_k \left\{ \exp\left(\mathrm{E}Z(C_k) + 1/2\sigma_z \right) - \mathrm{E}Y(B) \right\} = 0\end{aligned}$$

where $\beta = \{\beta_1, \beta_0, \sigma_z\}$. Based on aggregated demand in Section 4.1, the average-demand in the Texas region is given as $\mathrm{E}Y(B) = 206.5$.

The estimates $\tilde{\beta}$ in the hybrid likelihood given in Section 3.2 are obtained in two stages: first construct the profile likelihood in (11), and maximize it with respect to β. One of the key steps in the maximization is to estimate the Lagrange multiple τ. For a given β, Newton's method is used to solve for $\tau(\beta)$ in (10). Note the solution for $\tau(\beta)$ must satisfy $0 \leq w(C_k)/(\sum_k w(C_k) + \tau(\beta) g(x_k, \beta)) \leq 1$. For comparing our results with the traditional MLE of β without using the information from the higher hierarchy, we also calculate the traditional MLE $\hat{\beta}$ by simply maximizing the likelihood $\prod_{k=1}^{m} f(y_k; \beta, x_k)$.

The confidence intervals for β can be obtained by inverting the likelihood ratio test statistic (Pawitan (2001), page 35). Figure 6 shows the comparisons of the marginal profile likelihood of regression parameters constructed from the hybrid likelihood approach and

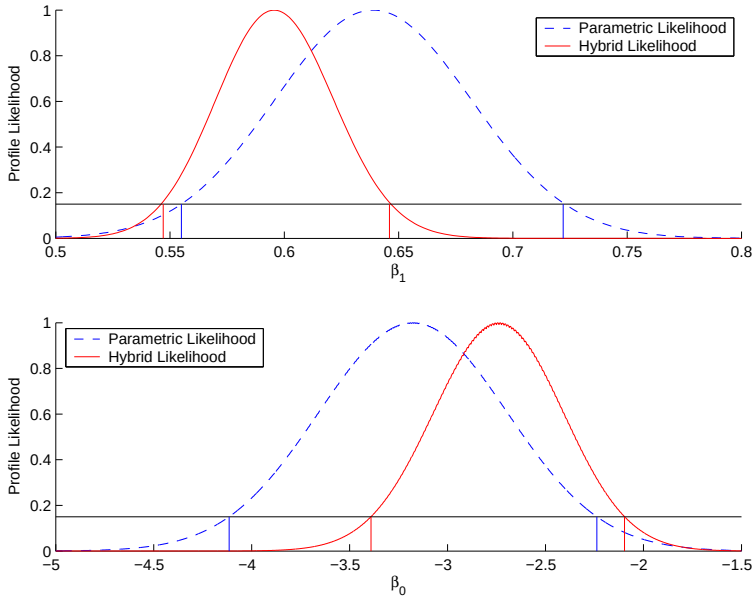

Figure 6: Confidence Intervals for β_1 and β_0

from the traditional likelihood approach. The horizontal lines on Figure 6 are drawn at a level such that the vertical lines intersecting the profile likelihood provide approximate 95% confidence intervals. The estimates and the confidence intervals for the traditional MLEs $\hat{\beta}_0$, $\hat{\beta}_1$ and hybrid MLEs $\tilde{\beta}_0$, $\tilde{\beta}_1$ are given in Table 1. Compared with traditional confidence intervals, the confidence intervals obtained from the hybrid approach are significantly narrower. This is due to the added information incorporated into the calculation of the likelihood function. See Qin & Lawless (1994) or Qin (2000) for the further studies on the finite-sample properties of parameter estimates by incorporating estimating equations as constraints.

Table 1: Estimates of Regression Parameters (β_1, β_0)

	Estimates	Confidence Intervals
$\tilde{\beta}_1$	0.596	(0.547, 0.646)
$\hat{\beta}_1$	0.639	(0.555, 0.722)
$\tilde{\beta}_0$	-2.744	(-3.391, -2.097)
$\hat{\beta}_0$	-3.178	(-4.115, -2.238)

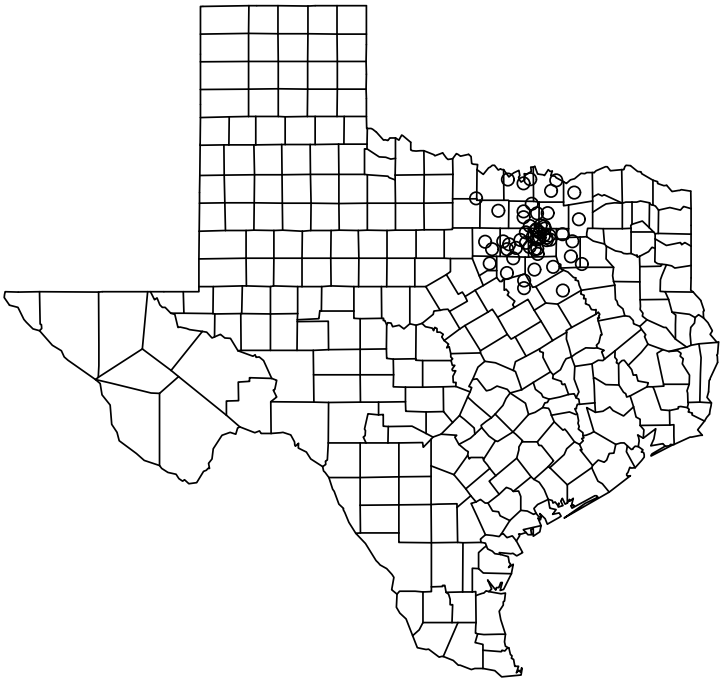

Figure 7: Location Data of stores in the Local District

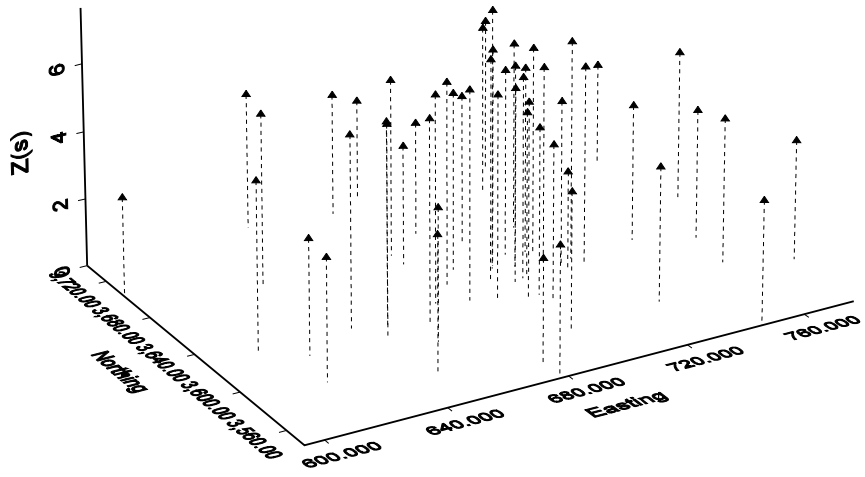

Figure 8: Perspective Plot of Transformed Demand and Location Data in the Local District

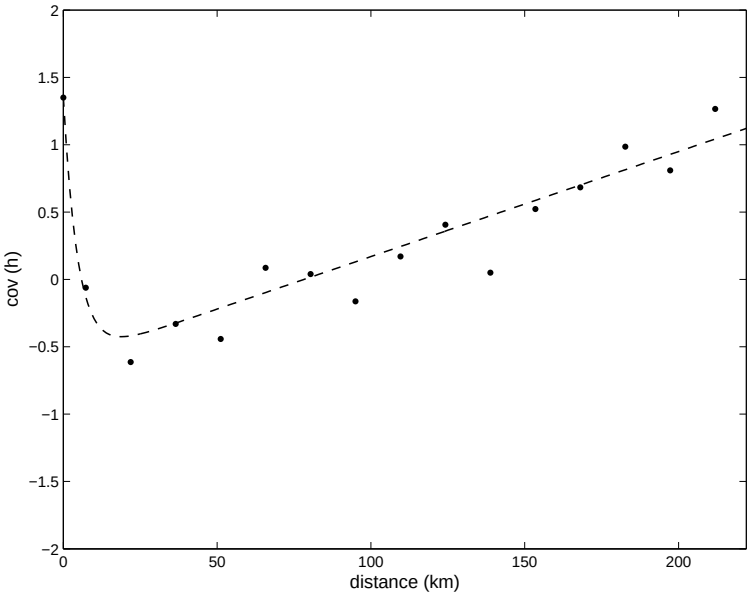

Figure 9: Sample Covariogram and Fitted Covariogram of Residuals in Local District

2.4.3 Local-Scale Log-Normal Kriging Using Linkage Information

Consider a local district $C \in B$ surrounding the city of Dallas as shown in Figure 7. Demands at existing stores are denoted as $Y(\boldsymbol{s}_l), l = 1, \ldots, L$ from $L = 55$ stores in this local district. Similar to the regional-study, the spatial analysis is performed in the transformed scale. Assume the local store demand can be modeled by a spatial process $Y(\boldsymbol{s})$, with $Z(\boldsymbol{s}) = \log(1+Y(\boldsymbol{s})) = x(\boldsymbol{s})'\beta+\delta(\boldsymbol{s})$, where β is the available coefficient information available from regional level, and $\delta(\boldsymbol{s})$ is a Gaussian Process with mean 0 and covariance $\boldsymbol{C}_\delta(\cdot)$. Figure 8 shows the three-dimensional scatter plot of the transformed demand data $Z(\boldsymbol{s})$ where the location coordinates $\boldsymbol{s} = (s_1, s_2)$ are in kilometers of the Universal Transverse Mercator (UTM) metric. After removing the mean effect, the variogram $\gamma_\delta(\boldsymbol{h})$ can be investigated for spatial dependence of $\delta(\cdot)$. Our analysis shows that there is no directional effect in spatial dependence, so we can regard the covariance functions as isotropic functions, i.e., $\boldsymbol{C}_\delta(\boldsymbol{h}) = \boldsymbol{C}_\delta(h)$, where $h = ||\boldsymbol{h}||$.

The covariance function is $\boldsymbol{C}_\delta(h) = \boldsymbol{C}_\delta(0) - \gamma_\delta(h)$. Due to the accuracy of the coordination systems, the stores within the same zip code have the same coordinates. The variance can be estimated by using the sample variance of store demands within the same zip code,

$\hat{\sigma}_\delta^2 = \hat{\boldsymbol{C}}_\delta(0) = 1.353$. The sample covariogram reveals interesting characteristics of the spatial dependence of the store demand. Figure 9 shows the isotropic empirical covariance function $\boldsymbol{C}_\delta(h)$ based on the data $\boldsymbol{\delta} = (\delta(\boldsymbol{s}_1), \ldots, \delta(\boldsymbol{s}_L))$. When the distance between two stores is small, the store demand has negative correlations due to competition in sales. Considering the stores are clustered around the Dallas metropolitan area, the similarity of store demands at suburban area contributes to the increasing correlations as h increases. This effect is different from the conventional monotone decreasing or continuous oscillating covariance functions (Cressie 1993). Based on the trend of empirical variogram, we construct the covariance function as an additive model,

$$\boldsymbol{C}_\delta(h) = \theta_1 \exp(-\theta_2 h) + \theta_3 h + \theta_4, \tag{18}$$

where $0 \leq h \leq h_{\max}$. In the first part of (18), the covariance function decreases exponentially to ensure the continuity of the covariance function. The upper bound $h_{\max}$ is to limit our focus within the local district, such that we do not need to extrapolate the spatial dependence structure outside of current district. According to Abrahamsen (1997), the function defined in (18) satisfies the conditions of positive definite functions, thus it is a valid covariance function. The parameters can be fitted using restricted maximum likelihood method (Cressie (1993), page 92) as, $\hat{\theta}_1 = 1.97$, $\hat{\theta}_2 = 0.21$, $\hat{\theta}_3 = 0.0078$, and $\hat{\theta}_4 = -0.61$, respectively. Based on the covariance function $\boldsymbol{C}_\delta(h)$, the $\delta_{opt}(\boldsymbol{s}_0)$ can be predicted using (13). Figure 10 shows the contour of the Kriging surface of residuals obtained at various locations in the local district. Due to the fast decrease of covariance function in short distance, the kriging surface is not smooth in small neighborhoods of existing observations. This kriging surface for the store demand can then calculated by adding the mean component $Z_{opt}(\boldsymbol{s}) = x(\boldsymbol{s})'\beta + \delta_{opt}(\boldsymbol{s})$, and then transformed to the original scale $Y_{opt}(\boldsymbol{s})$ using the bias-corrected formula (21) to predict the future store demand at any given location $\boldsymbol{s}$ within the local district.

2.5 Conclusion and Future Work

This article proposes multi-level spatial approximation models for combining data from various sources in different scales to support decisions at various levels. The information

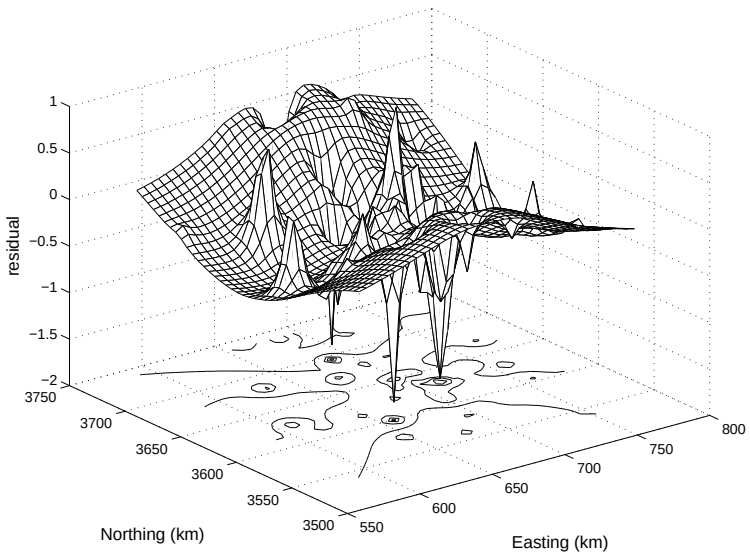

Figure 10: Kriging Surface of Residuals in Local District

from a higher hierarchy is structured in estimating equations serving as constraints in guiding the statistical inference at lower levels. These models can support strategic, tactical and operational decisions at different levels. Interesting logistical and spatial-prediction examples illustrate the proposed models and inference procedures.

The proposed methods are useful in various supply-chain, logistics and transportation decisions. For example, the constant-intensity CA model used in Dandamudi and Lu (2003) for supply-chain contract decisions involving logistics cost approximations can be extended to our models for more accurate cost-estimates. In the studies of Mangotra and Lu (2005) the spatial logistics model also plays a key role in deciding inventory levels coordinated between national- and regional-level DCs. These decisions will be affected greatly by the statistical modeling procedures and the associated estimation quality. Further studies are needed.

2.6 Appendix

Appendix A - Asymptotic Properties of the Parameter Estimates in Section 3.2

The Lagrange multiplier τ plays a key role in deriving the asymptotic properties of

the estimates $\hat{\beta}$. Similar to Qin and Lawless (1994), under proper regularity conditions, $\tau = O_p(n^{-1/2})$, for $\beta \in \{\beta : \|\beta - \beta_0\| \leq n^{-1/3}\}$. Thus, the asymptotic properties can be derived from using Taylor expansions of (10) and (12), in the neighborhood of the true value $\tau_0 = 0$ and β_0, respectively. Suppose $m/n \to a_0$ as $n, m \to \infty$, where $0 < a_0 \leq 1$. Then, we obtain the following theorem.

Theorem 1 *Let β_0 be the true value of the parameter β, where β and τ are given in (9). Assume that the regularity conditions on the normality of the MLE for typical parametric models as given in Lehmann (Chapter 6, 1998) and the regularity conditions specified in Qin and Lawless (1994) for estimating equations are satisfied (see the proof for details). For the hybrid likelihood given in Section 3.2, we have*

$$\sqrt{n}\begin{pmatrix} \hat{\beta} - \beta_0 \\ \hat{\tau} - 0 \end{pmatrix} \to N(0, \boldsymbol{U}),\ \boldsymbol{U} = \boldsymbol{S}\boldsymbol{V}\boldsymbol{S}^T,$$

and S,V are given as

$$\boldsymbol{S} = \begin{pmatrix} a_0\boldsymbol{J} & \dfrac{EW\partial g/\partial\beta}{EW} \\ -\dfrac{EW\partial g/\partial\beta'}{EW} & \dfrac{EWg^2}{E^2W} \end{pmatrix}^{-1},\ V = \begin{pmatrix} a_0\boldsymbol{J} & 0 \\ 0 & \dfrac{EWg^2}{E^2W} \end{pmatrix}$$

and $\boldsymbol{J} = E\left\{\left(\dfrac{\partial \log f(y_k; x_k, \beta)}{\partial\beta}\right)\left(\dfrac{\partial \log f(y_k; x_k, \beta)}{\partial\beta}\right)'\right\}$.

Proof: For notational convenience, let $w_k = w(C_k)$, $y_k = y(C_k)$. Define

$$\begin{aligned} Q_{1n}(\beta, \tau) &= \sum_{k=1}^{n} \frac{w_k g(x_k, \beta)}{w_k + n\tau g(x_k, \beta)}, \\ Q_{2n}(\beta, \tau) &= \sum_{k=1}^{n} \frac{w_k \tau}{w_k + n\tau g(x_k, \beta)} \frac{\partial}{\partial\beta} g(x_k, \beta), \\ Q_{3n}(\beta, \tau) &= \frac{1}{n}\sum_{k=1}^{m} \frac{\partial}{\partial\beta} \log f(y_k; x_k, \beta)), \end{aligned}$$

where Q_{2n} and Q_{3n} are the components of (12).

Using a Taylor expansion at $(\beta_0, \tau_0 = 0)$, we have

$$\begin{aligned} 0 &= Q_{1n}(\beta_0, 0) + \frac{\partial}{\partial\beta} Q_{1n}(\beta_0, 0)(\beta - \beta_0) + \frac{\partial}{\partial\tau} Q_{1n}(\beta_0, 0) \cdot \tau + o_p(n^{-1/2}) \\ 0 &= -Q_{2n}(\beta_0, 0) + Q_{3n}(\beta_0, 0) - \frac{\partial}{\partial\beta} Q_{2n}(\beta_0, 0)(\beta - \beta_0) + \frac{\partial}{\partial\beta} Q_{3n}(\beta_0, 0)(\beta - \beta_0) \\ &\quad - \frac{\partial}{\partial\tau} Q_{2n}(\beta_0, 0)\tau + o_p(n^{-1/2}). \end{aligned}$$

In addition,

$$\begin{aligned}
\frac{\partial Q_{1n}(\beta_0,0)}{\partial\beta} &= \sum_{k=1}^{n}\frac{w_k\frac{\partial g(x_k,\beta)}{\partial\beta}}{\sum_{k=1}^{n}w_k} \to \frac{\mathrm{E}\left(w\frac{\partial g}{\partial\beta}\right)}{\mathrm{E}(w)} \\
\frac{\partial Q_{1n}(\beta_0,0)}{\partial\tau} &= -\sum_{k=1}^{n}\frac{nw_i g^2(x_k,\beta)}{(\sum_{k=1}^{n}w_i)^2} \to \frac{\mathrm{E}(wg^2)}{\mathrm{E}^2(w)} \\
\frac{\partial Q_{2n}(\beta_0,0)}{\partial\beta} &= 0, \quad \frac{\partial Q_{2n}(\beta_0,0)}{\partial\tau} = \sum_{k=1}^{n}\frac{w_k\frac{\partial}{\partial\beta}g(x_k,\beta)}{\sum_{k=1}^{n}w_k} \to \frac{\mathrm{E}(w\frac{\partial g}{\partial\beta})}{\mathrm{E}(w)} \\
\frac{\partial Q_{3n}(\beta_0,0)}{\partial\beta} &= \frac{1}{n}\sum_{k=1}^{m}\frac{\partial^2}{\partial\beta\partial\beta'}\log f(y_k;x_k,\beta) \to a_0\mathrm{E}\left(\frac{\partial^2\log f(y_k;x_k,\beta_0)}{\partial\beta\partial\beta'}\right).
\end{aligned}$$

Thus,

$$\sqrt{n}\begin{pmatrix}\beta-\beta_0\\ \tau-0\end{pmatrix} = \begin{pmatrix}-\frac{\partial Q_{3n}}{\partial\beta} & \frac{\partial Q_{2n}}{\partial\tau}\\ -\frac{\partial Q_{1n}}{\partial\beta} & \frac{\partial Q_{1n}}{\partial\tau}\end{pmatrix}^{-1}_{(\beta_0,0)}\begin{pmatrix}\sqrt{n}Q_{3n}(\beta_0)\\ \sqrt{n}Q_{1n}(\beta_0,0)\end{pmatrix} + o_p(1) \to N(0,\boldsymbol{U}).$$

Appendix B - Derivation of Kriging Prediction for Log-Normal Observations

Combining the information from the block-averages at the higher hierarchy and individual observations at the lower hierarchy is difficult with non-Gaussian processes. Transformation techniques are used to solve this problem. We illustrate the Kriging procedure for transformed data using the log-normal case.

Let $Y(\boldsymbol{s})$ be a log-normal spatial process defined on C_0 so that the process $Z(\boldsymbol{s}) = \log(Y(\boldsymbol{s}))$ is a Gaussian process. Denote by $Z(C_0) = \log Y(C_0)$, and assume $Z(\boldsymbol{s}) = Z(C_0) + \delta_Z(\boldsymbol{s})$, where $\delta_Z(\boldsymbol{s})$ is a Gaussian process with mean zero and covariance function $\sigma_Z(\boldsymbol{u},\boldsymbol{v})$ for $\boldsymbol{s},\boldsymbol{u},\boldsymbol{v} \in C_0$. We first derive the optimal predictor on the log-scale, and then transform it back to the original scale. From Cressie (1993), we know that

$$\begin{aligned}
\mu_Y(\boldsymbol{s}) &= \mathrm{E}(Y(\boldsymbol{s})) = \exp\{\mu_Z(\boldsymbol{s}) + \sigma_Z(\boldsymbol{s},\boldsymbol{s})/2\}, \qquad (19)\\
\sigma_Y(\boldsymbol{u},\boldsymbol{v}) &= \mathrm{Cov}(Y(\boldsymbol{u}),Y(\boldsymbol{v})) = \mu_Z(\boldsymbol{u})\mu_Z(\boldsymbol{v})[\exp\{\sigma_Z(\boldsymbol{u},\boldsymbol{v})\} - 1].
\end{aligned}$$

Adjustments can be made to preserve the unbiasedness of the predictor. Based on $\mathrm{E}(Y(C_0))$ known from the regional-level study, $\mu_Z(C_0) = \mathrm{E}(Z(C_0)) = \log\mathrm{E}(Y(C_0)) -$

$\sigma_z(0)/2$. Then, $\mu_Z(C_0)$ is used in the following equation for the optimal estimator of $Z(\boldsymbol{s}_0)$ on the Z-scale.

$$\hat{Z}(\boldsymbol{s}_0) = \boldsymbol{c}_z' \Sigma_z^{-1} (\boldsymbol{Z} - \mu_z(C_0)\mathbf{1}') + \mu_z(C_0), \tag{20}$$

where $\boldsymbol{c}_z = (\sigma_z(\boldsymbol{s}_0, \boldsymbol{s}_1), \ldots, \sigma_z(\boldsymbol{s}_0, \boldsymbol{s}_L))$, $\boldsymbol{Z} = (Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_L))$, $\mathbf{1}$ is a unit column-vector and Σ_z is an $L \times L$ matrix whose (i, j)th element is $\sigma_z(\boldsymbol{s}_i, \boldsymbol{s}_j)$. Then, an unbiased predictor for $Y(\boldsymbol{s}_0)$ can be obtained by transforming $\hat{Z}(\boldsymbol{s}_0)$ back to Y-scale as follows:

$$\hat{Y}(\boldsymbol{s}_0) = \exp\{\hat{Z}(\boldsymbol{s}_0) + \sigma_z(\boldsymbol{s}_0, \boldsymbol{s}_0)/2 - \boldsymbol{c}_z' \Sigma_z^{-1} \boldsymbol{c}_z/2\}. \tag{21}$$

See Cressie (1993) for more discussions in log-normal kriging.

Table 2: Store Counts for the 48 States in the Continental United States

Acronym	Counts	Acronym	Counts
AL	85	AZ	46
AR	78	CA	306
CO	61	CT	63
DE	14	FL	166
GA	159	IA	150
ID	31	IL	275
IN	143	KS	94
KY	86	LA	87
MA	99	MD	74
ME	40	MI	210
MN	150	MO	154
MS	76	MT	48
NC	158	ND	43
NE	75	NH	33
NJ	125	NM	30
NV	18	NY	273
OH	238	OK	101
OR	62	PA	279
RI	11	SC	72
SD	48	TN	100
TX	349	UT	36
VA	126	VT	24
WA	80	WI	161
WV	55	WY	24

CHAPTER III

RELIABILITY MODELING IN SPATIALLY DISTRIBUTED LOGISTICS SYSTEM

3.1 Introduction

Logistics systems are designed to move goods, energy (e.g., electricity & gas), water, sewage, money, or information from origin to destination in a timely manner. A typical retail logistics system in a large-scale supply-chain has a few global distribution centers (DC), many local trans-shipment points, and thousands of stores. The service reliability of such a logistics system depends on the strategy of how stores are supported by a DC, and capability of the DC in replenishing products in a timely fashion in order to meet demands. Moreover, there are cases where uncertainties can disrupt logistics operations. For example, a labor dispute at Long Beach in 2002 shut down the port for 10 days, with an estimated cost in the range of $2 billion - $20 billion [23]. Recently, efforts have been made by the government & other stake-holders to insure the sea ports security to counter international terrorism, and to improve the security of the homeland. Research has been done [24] to plan the efficient inventory systems under possible occurrence of border disruptions. However, there are few analytical models available for supply chain reliability, despite the call for designing, and operating supply chains that are resilient to disruptions [25].

Past studies have examined problems in network reliability [26], system performance degradation, and workload re-routing for telecommunication, power, and transportation networks [27], [28]. Important characteristics for the performance of any telecommunication network are connectivity of communication, and capacity of transmittance [29]. Many articles concentrate on evaluating network reliability based on the probability of functioning in network links. For example, Varshney *et al.* [30] extended the performance measures with

a more general link-capacity model, allowing links to have multiple states with different capacities in each state. In telecommunication networks, especially packet-switched networks, past research has been devoted to designing robust networks, and maintaining high performance operational quality. When networks have serious outages or temporary disruptions, different sets of switch elements & interconnection links can be selected as alternatives to achieve fault tolerance. Some routing schemes have been proposed to optimize the network utilization, based on stochastic traffic flow dynamics & network topology. Many previous studies investigated these problems from different perspectives. For example, Yokohira *et al.* [31] proposed a fault tolerant network design based on a fixed routing scheme. Yang & Silvester [32] considered network fault tolerance problems where one switch can be replaced by another switch upon failure, thus showing that communication networks can be designed to be robust not only to link failures but also node failures. Mohorcic *et al.* [33] addressed adaptive routing problems to study the relative impacts of traffic load & propagation delay on the link cost.

A logistics network is distinguished from a telecommunication network by several characteristics. To handle uncertainty in demand, supply, and transportation in a logistics network, network nodes (DC & stores) need to coordinate to keep sufficient inventory to prevent stockout (running out of inventory). The cost of inventory includes interest on funds used to acquire goods kept in warehouses, storage space costs, warehouse labor, facility overhead, and paperwork for tracking & planning inventory levels. The telecommunications equivalent of inventory is a buffer space. While telecom nodes cannot hold spare inventory, they can buffer the data to send. While the main purpose of keeping inventory at logistics nodes is to prevent stockout due to fluctuations in demand & transportation time, a telecommunications network does not use buffer space to prevent stockout of data at users' facilities. In designing a logistics network & evaluating its performance, inventory cost is a tradeoff with transportation cost. The typical transportation costs are from equipment (e.g., truck rental), gas consumption, and labor. Unlike the high cost of the construction of telecommunication links, logistics networks usually rely on public roads. Conversely, while the cost of using the constructed network links is less in telecommunications, the

transportation cost on the logistics network roads is much higher.

Past studies [34] on road-transportation reliability addressed the travel time issues for trucks routing through connected stops. However, the size of the network in these studies is much smaller than the logistics systems considered in this article. Consequently, detailed probability-based reliability models have limited use in our supply-chain logistics systems. Also, in contrast to transportation networks, users of logistics networks often share the same goal (e.g., shipping products from various sources via a certain DC to stores). Thus, a more integrated solution can be proposed to coordinate network flows to minimize logistics costs, and to improve service levels.

Reliability studies of logistics service systems are also different from reliability research in product designs [83] where there are no DC, stores, or connectivity problems existing between them. The characteristics of a supply-chain's system reliability are discussed in only a few articles. For example, Bundschuh *et al.* [36] considered an integrated inbound supply chain with many potential suppliers for products. Suppliers are selected so that the supply-system remains reliable in case there is a low probability of failing to get needed supplies. Focusing on logistics cost instead of service reliability, Snyder & Daskin [25] studied situations where facility (e.g., DC) failures lead to excessive transportation costs, and they used mixed integer programming methods to minimize the increase of transportation costs under various failure scenarios. Thomas [37] proposed probability models for quantifying reliability of supply chains under contingency situations. However, he only considered simple probability models of supply meeting demand, $Pr(\text{supply} > \text{demand})$, based on assumed parametric distributions of supply & demand quantities. The complicated network issues between DC & stores were not discussed.

This article focuses on developing "first-cut" reliability models for large-size logistics systems to support strategic planning decisions such as DC location allocation & serving area optimization [38]. Coordinating inventory between DC & stores is a much more complicated issue, and is deferred to a subsequent paper. Our key contribution in this paper is to bring statistical spatial modeling techniques to approximate store location & demand data. Then, system reliability models can be built on these approximations for entertaining

various scenarios of DC location designs & DC capacity constraints. The following provides more details of our multi-level modeling, and reliability evaluation procedures in handling large-size networks.

Continuum approximation (CA) is a useful tool for modeling logistics systems for supporting strategic decisions in transportation & supply-chain management [39]. The CA approach first simplifies logistics networks by assuming that spatial locations (of stores, facilities, etc.) are modeled via a density function $\lambda(\boldsymbol{x})$ which describes the number of points per unit area as a function of position $\boldsymbol{x}$. This article extends the homogeneous Poisson intensity function to the nonhomogeneous case, which is capable of handling variations in store counts, and demands at different locations. As a result, DC allocation & serving areas can be determined more realistically. In our system reliability evaluation, the distance between a store & its nearest DC is one of the key factors for meeting a product's time-sensitive replenish requirement. Given a set of DC locations & capacities, reliability of the entire SC-logistics system is then an average of reliability over all store locations & demands based on approximation models. Our primary example has eight DC serving stores across the continental United States. Section 4.1 shows that when one of the DC breaks down, the entire system reliability could drop as much as 10.2%.

For tactical level decisions in choosing the most reliable & cost-effective routes, a similar approximation model is built on local store locations & demands. The reliability of a local logistics system depends more on whether certain routes are functioning or not. Thus, the experience gained from the study of telecommunication network reliability is utilized to develop reliability evaluation procedures at this level. In this paper, we consider DC at both global level & local level. To distinguish between these two cases, we denote a global distribution center by GDC, and a local distribution center by LDC.

Section 2 presents a general framework for modeling the reliability of global-level logistics networks, and evaluating the effects of degradation in the DC's capacity. Section 3 defines local-level logistics reliability measures. In both sections, well-known spatial smoothing methods are utilized to obtain a spatial intensity function for supporting CA approximation & reliability evaluation. Numerical illustrations with real-life data are provided in Section

4. Conclusion & future work are offered in Section 5.

3.2 Reliability Modeling for Global SC-Logistics Systems

Uncertainty is a primary concern in logistics planning. Often, safety stocks are utilized to accommodate uncertain demand or shipping delays. Logistics facilities such as DC are also subject to many types of risk. Snyder & Daskin [25] considered a situation where facilities fail time to time due to poor weather, labor disputes, or other factors, and applied the mixed integer programming method to find a contingent logistics solution. The focus of this article is to develop procedures for evaluating whether a large-size SC-logistics system is capable of handling certain contingencies for meeting product replenishment deadlines. Hereafter, the "stores" refer to stores in retail chains or dealers in automobile distribution networks.

To start, we present a rigorous definition of logistics service reliability. Consider a store located at $\boldsymbol{x}$, and assume it is served by a GDC $d(\boldsymbol{x})$ miles away. The speed of trucks transporting goods from the GDC to surrounding stores is modeled as a random variable V with a distribution function $F_V(\cdot)$. Then the service reliability at $\boldsymbol{x}$ for meeting the time deadline t_0 is defined & calculated as

$$\begin{aligned} r(\boldsymbol{x}) &= P(T \le t_0) = P[d(\boldsymbol{x})/V \le t_0] \\ &= P[V \ge d(\boldsymbol{x})/t_0] = 1 - F_V[d(\boldsymbol{x})/t_0]. \end{aligned} \tag{22}$$

The following example shows that the s-normal distribution with a reasonable choice of mean & standard deviation produces a proper reliability function. Figure 11 shows that the service reliability of a GDC to stores within a certain distance (e.g., 400 miles) is near one, i.e., the logistics system can meet the deadline without much trouble. For those stores far way from the serving GDC, the reliability starts to drop in a reversed-S shape curve, then reaches zero for stores quickly. These stores should be served by other GDC.

Example 1: Suppose V is s-normally distributed with mean 40 miles per hour, and standard deviation 10 miles per hour. Because the routes between the GDC & stores could include highways & local streets, this article sets the mean & standard deviation

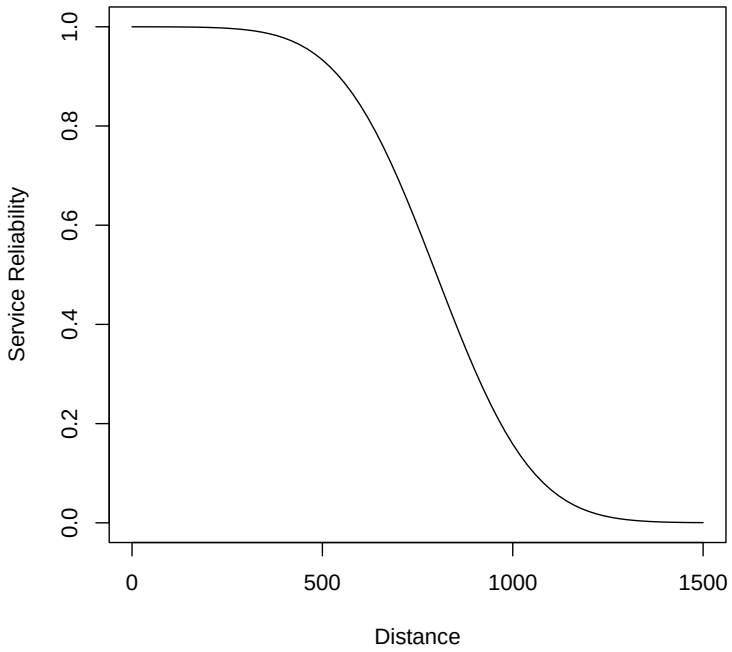

Figure 11: The service reliability $r(\boldsymbol{x})$ as a function of $d(\boldsymbol{x})$.

conservatively for accommodating possible varying factors such as road congestion, bad weather, and so on. Then the service reliability can be calculated as a function of distance $d(\boldsymbol{x})$, as shown in Figure 11. Notice that, within a 400-mile distance, the reliability is close to one. It drops to 80% when the distance is around 650 miles. The reliability is close to zero when the distance reaches 1300 miles from the GDC. One can imagine that when the deadline is more flexible (e.g., $t_0 = 24$ hours) the system is more reliable. In this case, the region of near 100% service reliability can be larger than the current region (about 400 mile radius distance) with $t_0 = 20$ hours.

3.2.1 Service Reliability for Large-size Logistics Systems

If one assigns specific stores to a specific DC, there will be an overwhelming number of possible combinations in store-to-DC assignments. This further complicates the calculation of system reliability. This section extends the reliability defined in (22) to many stores at locations $\boldsymbol{x} = (x_1, x_2)$ modeled by a smooth density function $\lambda(\boldsymbol{x})$. The demand for stores is also assumed to be a spatial function with mean $\mu(\boldsymbol{x})$. This makes the system-reliability calculation easier, and allows flexibility of handling many variations in the logistics system

with stores & DC organized differently. Our procedure explained in (23) below integrates the reliability of (22) over the serving region of a DC for assessing the service reliability in that region. Then we extend this idea to the whole region. The reliability for the entire system with many DC is then obtained

$$r_{system} = \int_A \lambda(\boldsymbol{u})\mu(\boldsymbol{u})r(\boldsymbol{u})d\boldsymbol{u} / \int_A \lambda(\boldsymbol{u})\mu(\boldsymbol{u})d\boldsymbol{u}. \tag{23}$$

Section 2.2 shows that this procedure allows us to easily handle problems with limited capacities at certain DC. If there are more DC added into the system (or some DC are dropped) to handle certain demand growth or contingency, the system reliability can always be evaluated with the same method as long as the serving regions are designated. It is possible to include reliability or other measures in the objective function (e.g., [38]) for deciding the serving region, and then calculate the resulted system reliability. However, up to this point, practitioners usually decide the serving region for a DC based on logistics, and facility cost concerns (e.g., [39]). Thus the scope of this article will be limited to evaluation of the reliability based on a given set of DC serving regions. That is, we do not discuss how to optimize DC locations & serving regions.

It is common in strategic planning of supply chain networks that exact store-location & demand information is often not available, or is projected based on aggregate planning or forecasting (see (29) & its description in Section 2.3 for details). Thus, this section follows the idea used in the Continuous Approximation (CA) approach for large-scale facility design problems by modeling store locations with a spatial Nonhomogeneous Poisson Process [43]. Its intensity function at the location $\boldsymbol{x} = (x_1, x_2)$ is $\lambda(\boldsymbol{x})$, which can be interpreted as an expected number of stores within unit area around $\boldsymbol{x}$. By using the CA approach, facility-location (e.g., DC location) allocation problems can be transformed to problems of finding the optimal size of DC service regions supporting the demands of stores in the whole region. Managers can change the intensity function (or the expected store counts in (29)) for projecting future situations when demand grows (or diminishes) in certain regions. Then the system reliability will be changed according to the change in the intensity function. This ability allows managers to evaluate if more DC are needed to improve its system reliability.

Next, let us illustrate the flexibility of the above setup in evaluating system reliability by considering that some DC might be shut down due to certain contingencies. Suppose there are n_I GDC serving the whole region A. If the GDC$_i$ is shut down (i.e., $\psi_i = 0$), the stores served by GDC$_i$ will be served by the next closest DC available. No matter how many GDC$_i$ are available, each store in the region will be served by a single GDC. Using the minimal distance concept, the serving regions of the GDC$_i$ are then defined. Thus, (23) can be used to evaluate the system reliability in this situation. See the algorithm presented at the end of this section for details of the calculation steps. See Example 3 in Section 4.1 for a numerical illustration.

Similar to (22), the service reliability $r^{\boldsymbol{\psi}}(\boldsymbol{x})$ under the scenario $\boldsymbol{\psi}$ is,

$$r^{\boldsymbol{\psi}}(\boldsymbol{x}; d^{\boldsymbol{\psi}}(\boldsymbol{x})) = 1 - F_V[d^{\boldsymbol{\psi}}(\boldsymbol{x})/t_0], \tag{24}$$

where $d^{\boldsymbol{\psi}}(\boldsymbol{x})$ is the distance of the $\boldsymbol{x}$ from the closest functioning GDC under the scenario $\boldsymbol{\psi}$. Then, the reliability under this scenario is $r^{\boldsymbol{\psi}}_{system}$, which is calculated by replacing $r(\boldsymbol{x})$ with $r^{\boldsymbol{\psi}}(\boldsymbol{x}; d^{\boldsymbol{\psi}}(\boldsymbol{x}))$ in (23). Finally, the expected service reliability of the logistics system under many possible scenarios is the weighted average of the reliability $r^{\boldsymbol{\psi}}_{system}$ with weights given as the probability of that scenario occurring. That is,

$$\mathrm{E}(r_{system}) = \sum_{\boldsymbol{\psi} \in \Psi} r^{\boldsymbol{\psi}}_{system} Pr(\boldsymbol{\psi}), \tag{25}$$

where $Pr(\boldsymbol{\psi}) = \prod_{i=1}^{n_I} (1 - p_i)^{\psi_i} p_i^{1-\psi_i}$.

Remark on Computing the Expected System Reliability: Enumerating all the possible scenarios may be cumbersome, however, (25) can be further simplified by evaluating the reliability at each $\boldsymbol{x}$ first, then integrating over the whole region,

$$\mathrm{E}(r_{system}) = \int_A \lambda(\boldsymbol{x}) \mu(\boldsymbol{x}) \mathrm{E}(r(\boldsymbol{x})) d\boldsymbol{x} / \int_A \lambda(\boldsymbol{u}) \mu(\boldsymbol{x}) d\boldsymbol{x}, \tag{26}$$

where the calculation of $\mathrm{E}(r(\boldsymbol{x}))$ can be carried out as follows:

- Consider each location $\boldsymbol{x}$. Denote GDC$_{(1)}$ as its serving DC.
- Order other GDC according to their relative distances from $\boldsymbol{x}$, and denote them as GDC$_{(2)}$, GDC$_{(3)}$, ..., GDC$_{(n_I)}$ with relative distances $d_{(2)}$, $d_{(3)}$, ..., $d_{(n_I)}$.

- Let $p_{(i)}$ be the failure probability for $\text{GDC}_{(i)}$. Denote $r(\boldsymbol{x}, d_{(i)})$ as the service reliability at $\boldsymbol{x}$ when it is served by $\text{GDC}_{(i)}$, when $\text{GDC}_{(1)}, \ldots, \text{GDC}_{(i-1)}$ have all failed.

- The $\text{E}(r(\boldsymbol{x}))$ inside the integration can be evaluated as

$$\begin{aligned} \text{E}(r(\boldsymbol{x})) &= \sum_{\boldsymbol{\psi}\in\Psi} Pr(\boldsymbol{\psi}) r^{\boldsymbol{\psi}}(\boldsymbol{x}) \\ &= (1-p_{(1)}) r(\boldsymbol{x}, d_{(1)}) + p_{(1)}(1-p_{(2)}) r(\boldsymbol{x}, d_{(2)}) + \ldots \\ &+ p_{(1)} \cdots p_{(n_I-1)}(1-p_{(n_I)}) r(\boldsymbol{x}, d_{(n_I)}). \end{aligned} \tag{27}$$

  For example, if the nearest GDC, namely $\text{GDC}_{(1)}$, is available, the reliability is the availability $(1-p_{(1)})$ multiplied by the reliability $r(\boldsymbol{x}, d_{(1)})$ based on the distance $d_{(1)}$ from store location $\boldsymbol{x}$ to $\text{GDC}_{(1)}$.

For practical reasons, a grid can be laid on the focused region such that (26) can be calculated by replacing integrals with summations as follows:

$$\text{E}(r_{system}) = \sum_i \sum_j \lambda_{ij}\mu_{ij}\text{E}(r^{ij}(\boldsymbol{x}))/\sum_i \sum_j \lambda_{ij}\mu_{ij},$$

where λ_{ij}, μ_{ij} is the total intensity & demand within the (i,j)th block in the grid, and $\text{E}(r^{ij}(\boldsymbol{x}))$ is the service reliability evaluated at the center of the grid, as denoted in (27).

3.2.2 Service Reliability in Logistics Systems with Capacitated DC

The reliability models discussed so far deal with problems of un-capacitated DC. That is, no matter how large the demand in a region has grown, the responsible DC can still handle its product replenishment, packing & loading tasks without losing any efficiency. This section extends the reliability models described in the previous section to the case where each DC can only handle limited demand. Denote that each GDC_i serving the region A_i has a fixed capacity ξ_i, $i = 1, \ldots, n_I$, and the union of all non-overlapped A_i is the whole region A.

Suppose that some contingent event (or demand growth) causes a "degradation" of GDC_i's capacity from ξ_i to ξ_i'. As a consequence, GDC_i can only handle demands from stores in a smaller subregion of A_i. The unserved stores need to go to other nearby GDC for service (if those GDC have available capacity to serve them). To implement this idea,

we revise the serving region for each GDC based on its capacity to meet demands in the nearest stores grid by grid, as described in the computing remarks above. When all the stores in the original region assigned to a particular GDC are served, and if this GDC still has extra capacity to serve other stores, the reliability computing algorithm will pick up the unserved stores in grids nearest to this GDC. All capacity at the GDC will be used; if there are still some stores unserved, the system reliability will drop. In this case, managers can consider adding capacity to certain GDC, or adding some temporary GDC. See Example 4 in Section 4.1 for numerical illustrations.

Remark: This idea is simple, but might not be optimal in the sense of maximizing system reliability. For example, consider two neighboring DC, where DC_1 has available capacity for handling the demands from one unserved store at location a that was originally served by DC_2. Suppose that the distance from this store to DC_2 is $d_{2a} = 5$ miles, and its distance to DC_1 is $d_{1a} = 8$ miles. According to the idea described above, this store should be served by DC_1. However, suppose that there is another store (at location b) originally served by DC_2 with equal demands that is $d_{2b} = d_{2a} + 1 = 6$ miles from DC_2, and is $d_{1b} = d_{1a} + 10 = 18$ miles to DC_1. Although the store at location b is slightly farther away from DC_2, its distance to DC_1 is much farther away than that from the store at location a. If we assign DC_2 to serve store b, the "saved" transportation distance (9 miles) gained by using DC_1 to serve store a will increase the system reliability, which is a monotone function of travelling distance. That is, we should look for the case where the store gains the most savings in travelling distance, but not the nearest store. Thus the algorithm should first calculate the differences for stores travelling to two DC, then choose the case saving the most. For example, the difference for store a to DC_1 & DC_2 is $8 - 5 = 3$ miles. Similarly, this difference for store b is $18 - 6 = 12$ miles, which is a much greater distance. Thus, store b should be served by DC_2, while store a should be served by DC_1.

This modified idea only works for equal demands in these two stores. The situation for unequal demands becomes even more complicated, especially in pooling demands from a few stores together to match the demand for another store to calculate the total distance

saved. Because this article represents initial research in introducing service-reliability in large-size supply-chain logistics systems, the scope of such an in-depth study is beyond the mission of this article. We will leave it to future research.

3.2.3 Global-level Spatial Density Modeling

In logistics planning at the corporate level, many strategic plans utilize data projected in aggregated forms. For example, the manager who handles marketing departments might forecast a total demand for many stores in a region, e.g., the state of Georgia or the South Eastern region. This section employs a commonly used smoothing technique for estimating the intensity function $\lambda(\boldsymbol{x})$ described in Section 2.1 based on constraints from the total store counts in a region set by the marketing manager. For simplicity, we assume all stores have the same expected demand μ_0.

As an example, consider the state B_j $(j = 1, \ldots, n_J)$ as a region where the marketing manager has some aggregated store counts. Figure 12 illustrates these aggregated counts for states in the continental U.S. (region A), where darker colors indicate larger counts. Because the State of Texas has the largest number of stores in this data set, its color is the darkest. See Table 2 for details of the data. Tobler [40] constructed an estimator $\hat{\lambda}(\boldsymbol{x})$ to minimize the following smoothness function (in second derivatives),

$$\iint_A \left[\left(\frac{\partial^2 \lambda(\boldsymbol{x})}{\partial x_1^2} \right) + \left(\frac{\partial^2 \lambda(\boldsymbol{x})}{\partial x_2^2} \right) \right]^2 dx_1 dx_2 \tag{28}$$

subject to constraints $\lambda(\boldsymbol{x}) \geqslant 0$, and

$$\int_{B_j} \lambda(\boldsymbol{x}) d\boldsymbol{x} = \mathrm{E}(W(B_j)), \text{ for } j = 1, \ldots, n_J, \tag{29}$$

where $\mathrm{E}(W(B_j))$ is the expectation of projected store counts in each subregion B_j. The procedure is a nonparametric smoothing method making interpolations of given data. The assumption is that there exists a density function which has a non-negative, finite value for every location in the domain. For the logistics data we are working on (see Section 4), this assumption will hold. Moreover, because we only consider stores within the continental U.S., the spatial density at the boundary is fixed at zero. The existence & uniqueness for

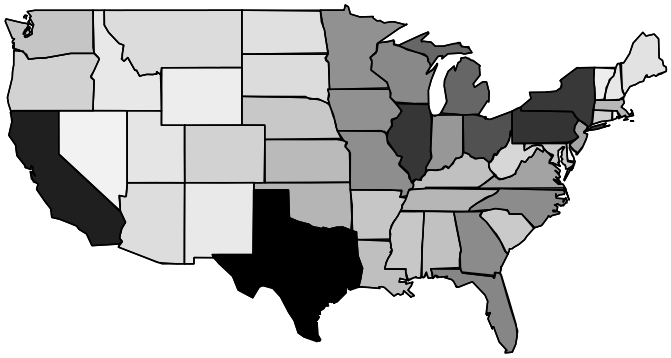

Figure 12: The counts $W(B_j)$ of stores for each B_j in the continental U.S.

this constrained optimization problem has been shown in [40]. Figure 15 in Section 4.1 demonstrates this procedure & the fitted smooth intensity function.

3.3 Reliability Evaluation of Local Logistics Systems

In local logistics operations, one needs to conduct more detailed decision analysis such as which local DC & routes to choose. The time constraints become more stringent. In this situation, local distribution centers (LDC) or transit points are closer to stores, thus being able to provide more timely service. In strategic decisions, failure of a GDC is a primary concern. However, at the local-level, one is more concerned about weather conditions, road traffic, or other factors that might affect tactical decisions on truck-routing & product-inventory. In this section, we assume that the transportation links are subject to failures, and contingent events may disrupt specific routes connecting the GDC, LDC & stores. To keep the study focused, we do not discuss inventory issues. See Dandamudi & Lu [38] for such studies.

For illustrative purpose, let us consider the local logistics network shown in Figure 13 where the "Origin" is a GDC. There are four local distribution centers (denoted by LDC_k, $k = 1, 2, 3, 4$) serving as trans-shipment points for aggregating & mixing products from different sources. Denote the "long-haul" transportation lines (called "links") between GDC & LDC as e_l, $l = 1, \ldots, n_L$, as illustrated in Figure 13 with $n_L = 5$. Thus we model the local logistics system as a stochastic network, where the links e_l have independent failure

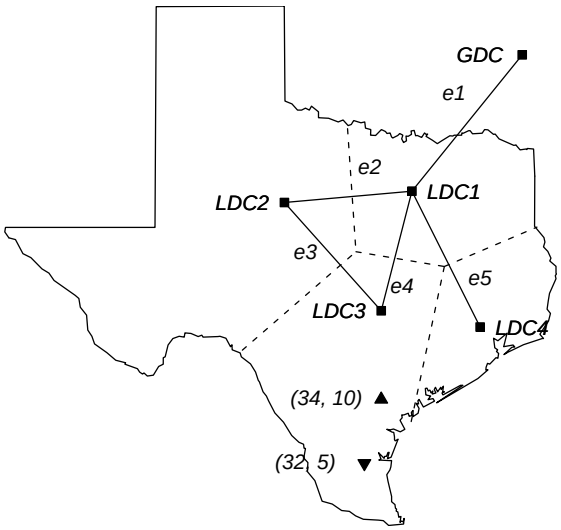

Figure 13: LDC locations & service regions in the state of Texas.

probability p_{e_l}, $l = 1, \ldots, n_L$.

After products are transported to LDC_k through long-haul links (e.g., interstate highways or railways), trucks will deliver the products to stores at $\boldsymbol{x}$ through local routes. Compared with local transportation speed, the long-haul transportation speed usually has less variation due to more stable traffic & road conditions. For simplicity, we assume the long-haul speed to be deterministic, which is denoted by v_{DC}. The local transportation speed is denoted by a random variable V with distribution function $F_V(\cdot)$.

Before defining the local-level service reliability $r_{system,L}$, we introduce several definition & notations used in telecommunication network reliability. First, a *minimal path* [41] is a minimal set of links whose simultaneous functioning ensures the functionality of a local logistics network. Define the *shortest path* between origin (GDC) & destination (store) to be the available minimal path which takes the least amount of time for minimizing the transportation cost. In our model, we assume the flow of products will always go through the shortest path available to maximize the service reliability of a region.

Example 2: Let's focus on the state of Texas with four LDC receiving products from a GDC, and shipping them to surrounding stores as shown in Figure 13. The link lengths (in miles) are calculated as $(l_{e_1}, l_{e_2}, l_{e_3}, l_{e_4}, l_{e_5}) = (500, 296, 300, 350, 356)$. To keep this

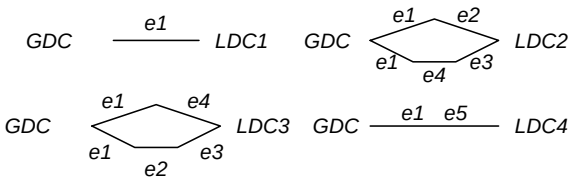

Figure 14: Minimal path sets for node pairs.

presentation short, assume every store has the same demand (μ_0) during the unit time period.

Figure 14 summarize all the minimum paths $\text{MP}_{k(l)}$ linking the four LDC with the GDC. For example, consider LDC_1 with only one minimal path, e_1, connecting the GDC. For LDC_2, it has 2 ($=g_2$) minimum paths, (e_1, e_2), and (e_1, e_4, e_3) from the origin. According to the distance calculation, the minimum path $MP_{2(1)} = (e_1, e_2)$ is the shortest path. The other path $MP_{2(2)} = (e_1, e_4, e_3)$ will be used when the critical link such as e_2 has failed. Similarly, LDC_3 has two minimum paths, $MP_{3(1)} = (e_1, e_4)$, and $MP_{3(2)} = (e_1, e_2, e_3)$, where $MP_{3(1)}$ is the shortest path here. Finally, LDC_4 has one minimum path, $MP_{4(1)} = (e_1, e_5)$.

The expected system reliability can be obtained as a weighted average of the reliability under scenario $s_{k(l)}$. For example, the scenario $s_{1(1)}$ means that products are delivered from the GDC to LDC_1 through link e_1. Thus, the probability of this scenario is $Pr(s_{1(1)}) = (1-p_{e_1})$. Following this idea, the probability of routing from $MP_{3(1)} = (e_1, e_4)$ is $Pr(s_{3(1)}) = (1-p_{e_1})(1-p_{e_4})$ for both routes e_1 & e_4 to function simultaneously. On the other hand, considering another minimum path of LDC_3, $MP_{3(2)} = (e_1, e_2, e_3)$. When e_4 fails, products will be routed through $MP_{3(2)}$ rather than $MP_{3(1)}$. Thus the probability in this special scenario $s_{3(2)}$ is $Pr(s_{3(2)}) = (1-p_{e_1})(1-p_{e_2})(1-p_{e_3})p_{e_4}$. By applying this idea, we can calculate all the other scenario-probabilities as follows: $Pr(s_{2(1)}) = (1-p_{e_1})(1-p_{e_2})$, $Pr(s_{2(2)}) = (1-p_{e_1})(1-p_{e_3})(1-p_{e_4})p_{e_2}$, $Pr(s_{4(1)}) = (1-p_{e_1})(1-p_{e_5})$. These probabilities will be used in Section 4.2 for evaluating the system reliability.

Under scenario $\boldsymbol{s}$, the long-haul transportation time $t_{ok}^{\boldsymbol{s}}$ is equal to the distance $d_{ok}^{\boldsymbol{s}}$ between O to LDC_k divided by the deterministic long-haul travelling speed v_{DC}. Denote

the distance between the LDC_k & its serving store at $\boldsymbol{x}$ as $d_k(\boldsymbol{x})$. Then the transportation time between the LDC & the store is $d_k(\boldsymbol{x})/V$. Note that, for simplicity, the local & short-distance routes between the LDC & stores are assumed to be always functioning. This implies that the total transportation time from the origin to the store is $T = d_k(\boldsymbol{x})/V + t^{\boldsymbol{s}}_{ok}$. Thus, the service reliability for a store at $\boldsymbol{x}$ is $r^{\boldsymbol{s}}(\boldsymbol{x}) = Pr(T \le t_0) = 1 - F_V[d_k(\boldsymbol{x})/(t_0 - t^{\boldsymbol{s}}_{ok})]$. Apply the store location & demand smoothing idea described in Section 2. The total service reliability under scenario $\boldsymbol{s}$ for the entire region B_0 (e.g., the state of Texas) is

$$r^{\boldsymbol{s}}_{system,L} = \int_{B_0} \lambda(\boldsymbol{x})\mu(\boldsymbol{x})r^{\boldsymbol{s}}(\boldsymbol{x})d\boldsymbol{x} / \int_{B_0} \lambda(\boldsymbol{x})\mu(\boldsymbol{x})d\boldsymbol{x}. \tag{30}$$

Consider all possible scenarios $\boldsymbol{s}$ with different occurrence probability $Pr(\boldsymbol{s})$. The expected system reliability for this region is then

$$\text{E}(r_{system,L}) = \sum_{\boldsymbol{s}\in\mathcal{S}} Pr(\boldsymbol{s}) r^{\boldsymbol{s}}_{system,L} = \int_{B_0} \lambda(\boldsymbol{x})\mu(\boldsymbol{x})\text{E}(r(\boldsymbol{x}))d\boldsymbol{x} / \int_{B_0} \lambda(\boldsymbol{x})\mu(\boldsymbol{x})d\boldsymbol{x}. \tag{31}$$

In the local regions, suppose a snapshot of detailed store locations $\boldsymbol{x}_m = (x_{1m}, x_{2m})$ $(m = 1, \ldots, n_M)$ is available, so we can use the following well-known kernel smoothing technique [42] to estimate the spatial intensity function at point $\boldsymbol{u} = (u_1, u_2)$,

$$\hat{\lambda}(\boldsymbol{u}) = \frac{1}{n_M} \sum_{m=1}^{n_M} \omega(u_1 - x_{1m}; h_1)\omega(u_2 - x_{2m}; h_2),$$

where the symmetric univariate density function (Gaussian) satisfies $\int \omega(\boldsymbol{u})d\boldsymbol{u} = 1$. The specification of the bandwidth $\boldsymbol{h} = (h_1, h_2)$ is important for practical implementation; however, it is beyond the scope of this paper to present the details how to select $\boldsymbol{h}$. Thus, we will use the following well-known estimate ([44], page 60) in our implementation in Section 4.2,

$$\hat{\boldsymbol{h}} = \left\{ 8\pi^{1/2} \int \omega(u)^2 du / [3n_M (\int u^2 \omega(u)du)^2] \right\}^{1/5} \hat{\boldsymbol{\sigma}}, \tag{32}$$

where $\hat{\boldsymbol{\sigma}}$ is an estimate of standard deviation of store locations. This estimator is known as the *optimal* bandwidth estimator to minimize the *mean integrated squared error* when the store locations are s-normally distributed. See [42] for a discussion on computation issues.

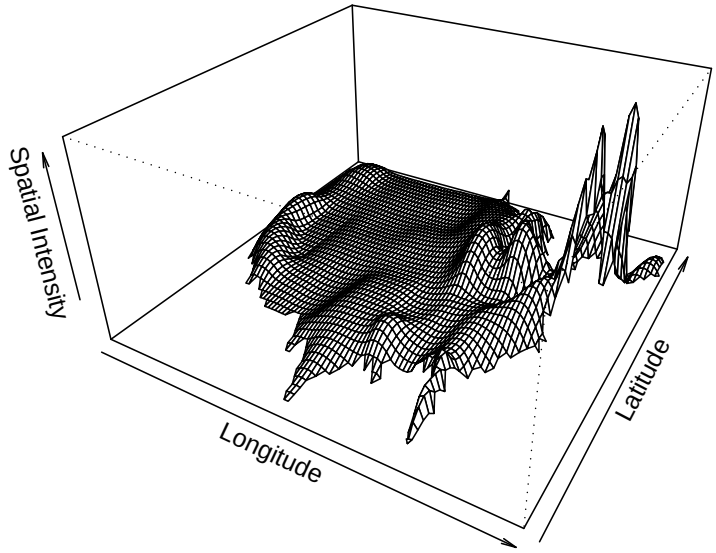

Figure 15: Estimated spatial density function in global level.

3.4 Applications of System Reliability Measures

3.4.1 Global-level Modeling & Reliability Analysis

The data used for illustrating the proposed reliability measures are contributed from a major U.S. automobile manufacturing company. The stores in this case represent car dealers. The store counts per state are provided in an aggregated form as shown in Table 2 & Figure 12. This company has eight global DC serving the whole U.S. See Figure 16 for their locations. Based on the count data $W(B_j)$, a smooth function $\hat{\lambda}(\boldsymbol{x})$ is obtained using the constrained interpolation method discussed in Section 2.3, where a 100×50 grid is overlaid on the U.S. map. All the computations are implemented using R [45]. Figure 15 presents the store intensity function. Note that the New York & Boston city areas have the highest intensities contributed from the greatest number of stores in the relatively small regions. Texas has the largest store count, but the size of the state is relatively large. Thus its store intensity is not as large as the New York & Boston areas.

Example 3: In this example, we calculate system reliability under the scenario of a single DC shutdown, and possible multiple DC shutdowns. Let V & the delivery deadline be the same as specified in Example 1. When all GDC are available, i.e., $\boldsymbol{\psi}_0 = (1, \ldots, 1)$, the

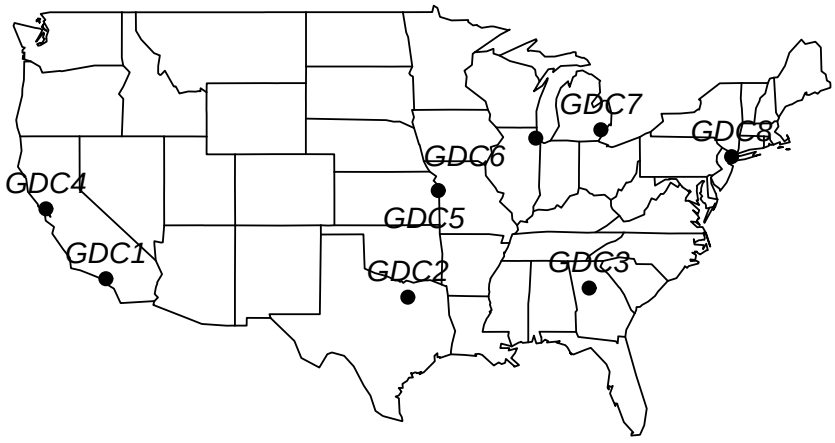

Figure 16: Locations of the 8 GDC in the continental U. S.

original system reliability is calculated as $r^{\psi_0} = 0.926$, using (23).

When only GDC_j shuts down, i.e., $\boldsymbol{\psi} = (1, \ldots, 0_j, \ldots, 1)$ & $p_j = 1$, (23) & (27) gives the system reliability as reported in Table 3. Notice that GDC_8 near Boston is the only GDC in the North-Eastern region with a lot of demands. Its shutdown should greatly affect the system reliability. According to our calculation given in Table 3, its shutdown does have the largest impact on system reliability, a 10.2% drop in the system reliability. The GDC_3 near Atlanta has the second largest impact with about a 7.7% drop in the system reliability. On the other hand, GDC_7 near Detroit has the least effect (0.4%) on the system reliability, because GDC_6 located at Chicago may serve as a backup.

Next, consider the case with multiple GDC shutdowns. For an illustration, suppose GDC_1 & GDC_4 shut down simultaneously, which means that there is no GDC in the West, including the highly populated California regions. Our calculation shows that the system reliability declines to 0.834, which is about 9.9%, slightly smaller than the effect of GDC loss near the Boston area. Consider another extreme case, where three GDC in the North, GDC_6, GDC_7, and GDC_8, are all shut down simultaneously. Then the system reliability reduces to 0.698, which is about a 24.6% drop from the original system reliability. When the probability of GDC shutdowns is available, we can evaluate the expected system reliability. For example, if we assign the probability of every GDC being shut down as 0.1 ($= p_j$), the

Table 3: The system reliability after a particular GDC fails

GDC	1	2	3	4	5	6	7	8
r^{ψ}	0.914	0.890	0.855	0.907	0.903	0.917	0.922	0.832

system reliability is 0.896, which is about a 3.2% drop from the original system reliability.

Remark: Most of the existing CA applications in the logistics-planning literature assume a constant intensity function (e.g., [39]); thus the spatial density issues addressed in this article have not been considered. The evaluation of service reliability may therefore not be accurate due to the mis-representation of store information. For a comparison purpose, the following provides a few reliability calculations based on this constant intensity model. In the case where every GDC is available, the service reliability for the whole region becomes $r_{system} = 0.750$, which is much lower than the 0.926 system reliability calculated from the spatial model. One can expect that when there are clusters of stores near a DC, the reliability will be higher. This is what the spatial model represents. When the stores are assumed to be uniformly distributed, the distance from stores to DC will be larger, thus the system reliability will be lower.

When only the ith GDC fails, $i = 1, 2, \ldots, 8$, the service reliability presented in Table 3 in the constant intensity case become lower, and their values are 0.711, 0.699, 0.708, 0.715, 0.708, 0.745, 0.749, 0.720. More importantly, the order of the reliability are changed. For example, shutting down GDC_1 has a larger impact to the system reliability than shutting down GDC_8. Their reliability are 0.914 & 0.832, respectively, as shown in Table 3. However, when the constant intensity model is used, their values become 0.711 & 0.720. An incorrect reliability evaluation could result in less attention toward maintaining GDC_1's functionality.

Example 4: This example considers the capacitated GDC, and evaluates the deterioration of system reliability when one GDC's capacity is degraded gradually due to demand growth, or certain types of contingencies, e.g., a labor dispute. Assume all stores have homogeneous expected demand μ_0, then the total expected store demand μ_j served by GDC_j is then $\mu_j = \int_{A_j} \mu_0 \lambda(\boldsymbol{x}) d\boldsymbol{x}$. Suppose each GDC is built with capacity $\xi_j = 1.2 \times \mu_j$. Thus,

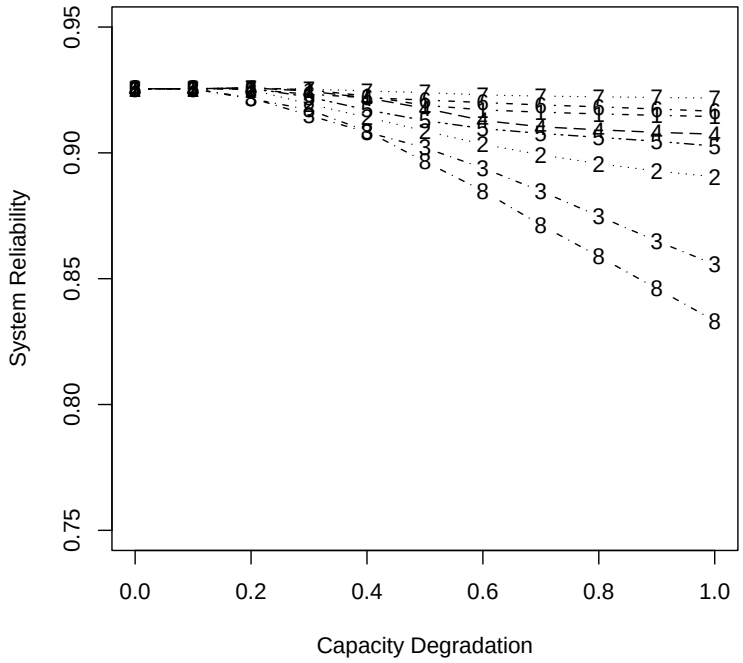

Figure 17: The reliability degradation paths for each GDC when its capacity decreases. x-axis shows the available capacity.

neighboring GDC can share the load to counteract the degradation of one GDC's capacity. Figure 8 shows the reliability degradation path for each GDC when its available capacity decreases from 100% (no capacity change) to 0 (complete shutdown of the GDC).

According to these path patterns, the service reliability in GDC_8 & GDC_3 degrade the most; much larger than the others, especially after about a 50-60% capacity degradation in these two GDC. The reason is that those DC are further away from neighboring DC, and have relatively larger store demands. On the other hand, the capacity degradation of RDC_6 & RDC_7 serving the Chicago & Detroit regions has little impact on the system reliability due to their closer distance from each other.

3.4.2 Local-level Modeling & Reliability Analysis

Figure 18 shows store locations in Texas. Using the s-normal kernel spatial smoothing method described in Section 3, the bandwidth estimator is calculated as $\boldsymbol{h} = (h_1, h_2) = (0.604, 0.522)$. In this example, we lay a 50×50 grid on this region, and obtain the estimated smooth spatial density of store locations as illustrated in Figure 19.

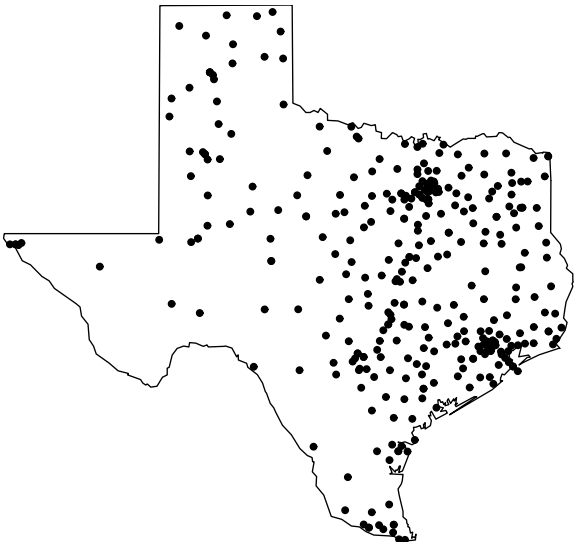

Figure 18: A snapshot of store locations in the state of Texas.

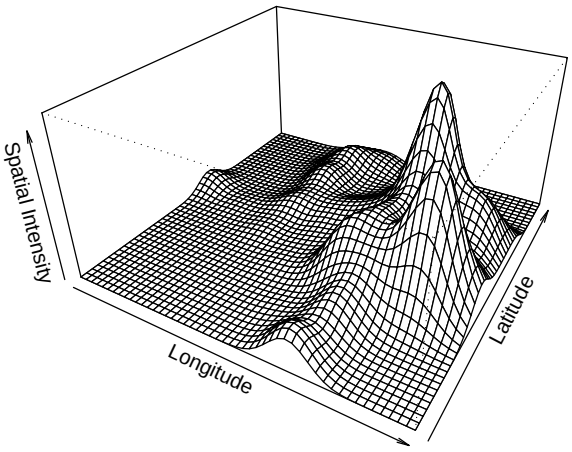

Figure 19: Estimated spatial density function in local level .

Example 5: By using the minimum shortest paths, and many scenarios described in Example 2, we can calculate the probability of routing through each minimal path & its corresponding service reliability for stores. For example, consider a grid block of stores represented by the block center $(34, 10)$. Its distance to LDC_3 & LDC_5 are 186.3 & 253.36 miles, respectively. Thus, based on transportation cost & reliability concerns, stores located at this block will be served by LDC_3. Assume all the links have same failure probability $p_{e_j} = 0.05$. The probability of the scenario for routing through the first minimal path $\text{MP}_{3(1)} = (e_1, e_4)$ from GDC to LDC_3 is $Pr(s_{3(1)}) = 0.90$.

Next, we use (30) to calculate the service reliability. Consider the deadline $t_0 = 24$ hours from GDC to a store in the block $(34, 10)$ (represented by a triangle pointing upward in Figure 13). The transportation time on the $MP_{3(1)}$ with link distances $e_1 = 500$ & $e_4 = 350$ miles is $t_{o3}^{s_{3(1)}} = (l_{e_1} + l_{e_4})/v_{DC} \approx 14.17$ hours, where $v_{DC} = 60$ miles per hour is the deterministic transportation speed between GDC & LDC. Then the service reliability is

$$r^{s_{3(1)}}(\boldsymbol{x}) = Pr(d_3(\boldsymbol{x})/V + 14.17 \leq 24) = 1 - F_V(d_3(\boldsymbol{x})/9.83) = 0.98,$$

where the location in this example is $\boldsymbol{x} = (34, 10)$, and the distribution of V is s-normal with mean & standard deviation equal to 40 & 10 miles per hour, respectively.

Similarly, the probability of routing through the second minimal path $\text{MP}_{3(2)} = (e_1, e_2, e_3)$ is $Pr(s_{3(2)}) = 0.95 \times 0.95 \times 0.95 \times 0.05 = 0.04$, and it will take $(l_{e_1} + l_{e_2} + l_{e_3})/v_{DC} \approx 18.27$ hours to pass the long-haul distance. Under this scenario, the service reliability is then

$$r^{s_{3(2)}}(\boldsymbol{x}) = Pr(d_3(\boldsymbol{x})/V + 18.27 \leq 24) = 1 - F_V(d_3(\boldsymbol{x})/5.73) = 0.77.$$

The service reliability for serving the store at $\boldsymbol{x} = (34, 10)$ is

$$\text{E}(r(\boldsymbol{x})) = Pr(s_{3(1)})r^{s_{3(1)}}(\boldsymbol{x}) + Pr(s_{3(2)})r^{s_{3(2)}}(\boldsymbol{x}) = 0.90 \times 0.98 + 0.04 \times 0.77 = 0.92.$$

Consider another grid block $\boldsymbol{x}' = (32, 5)$, which is shown as a triangle pointing downward in Figure 13. Its distance (283 miles) to LDC_3 is shorter than distances from other LDC. Thus, LDC_3 will serve this block. Because the distance from LDC_3 to this location is much larger than the distance (186.3 miles) from the same LDC to location $\boldsymbol{x} = (34, 10)$, the

following shows that the service reliability to this location is lower (0.79 versus 0.92) due to the larger distance. Using the same procedure as illustrated in the paragraph above, we can calculate the service reliability routing through its first & second minimal paths as 0.87 & 0.17, respectively. Thus the service reliability evaluated at this block $\boldsymbol{x}'$ is $\mathrm{E}(r(\boldsymbol{x}')) = 0.90 \cdot 0.87 + 0.04 \cdot 0.17 = 0.79$.

After enumerating all the stores at grid-locations within the service region of LDC_3, we can calculate the local-level system reliability using (30) as $\mathrm{E}(r_{system,L3}) = 0.87$. By going through all detailed calculations for all LDC & minimum paths, the total local service reliability can be calculated as $\mathrm{E}(r_{system,L}) = \int_{B_0} \mathrm{E}(r(\boldsymbol{x}))\lambda(\boldsymbol{x})\mu(\boldsymbol{x})d\boldsymbol{x} / \int_{B_0} \lambda(\boldsymbol{x})\mu(\boldsymbol{x})d\boldsymbol{x} = 0.90$.

Now, let's consider a few other examples with different link failure probabilities. Choose $p_{e_1} = 0.15$ while keeping other probabilities the same as above. Then the total system reliability drops to 0.80. If $p_{e_5} = 0.15$, and keeping all others the same, the service reliability will drop to 0.87. Similarly, if we change p_{e_3} to 0.15 then $r_{system,L}$ becomes 0.89. This example shows the relative importance of these links in affecting local system service reliability. The link e_1 is involved in all the LDC service-regions, so it is most critical. Thus its failure will lead to a larger drop in the system reliability. The link e_5 is slightly more important than the link e_3, which can be regarded as a backup link between LDC_2 & LDC_3.

3.5 Conclusion & Future Work

This article employs a spatial approximation method to characterize patterns of store locations (and demands) in a large-size supply-chain logistics system. Based on the approximation models, different reliability evaluation schemes are proposed for logistics planning at global & local levels. Service degradation patterns are examined to understand their relationship with DC's capacity limitation.

When we consider the transportation, facility, labor, and other costs of DC operations, the optimal locations of DC & the re-routing strategy become more complicated. How to design a system with minimal cost, and robust against possible capacity degradations (or fully shut down) under unexpected contingencies, is a challenging problem. In addition,

if the connectivity between the DC & stores is not fully guaranteed, the system reliability evaluation & robust system design issues become even more complex. This article sets the foundation of research in these directions, and shows the potential of research in the service sector focusing on logistics-network reliability evaluation issues.

CHAPTER IV

DETECTION AND ESTIMATION OF A MIXTURE IN A POWER LAW PROCESS FOR A REPAIRABLE SYSTEM

4.1 Introduction

For a repairable system, it is crucial to understand how the system's failure rate changes over usage time so that it can be taken off test before repairs become too frequent and too costly to the system operation. In many industries, the challenge is in finding which systems are susceptible to frequent failures and which ones are not. For complex systems that are susceptible to competing risks (i.e., risks of failure from different sources) and wear due to usage, it is not uncommon to end up with a mixture of good (high quality) systems with a small fraction of defective ones.

We refer to these good systems as "conforming" because the quality might be measured not only in terms of reliability (e.g., time on test) but in other aspects having to do with system operation. The "nonconforming" systems will exhibit shorter operation time between repairs, and unlike a repair to the non-defective conforming systems, these repairs can include different failure modes that are seemingly unrelated.

In the automotive industry, for example, the small proportion of new cars that make repeat trips to the repair shop are called *lemons*, and several states have adopted consumer protection rights ("lemon laws") that will force the manufacturer to replace the defective product with no cost to the consumer. There is an industry of law practices just for lemon law cases, as pointed out in Lehto (2000) and Megna (2003).

By treating the defective products as a (contaminated) sub-population, the time to failure of a new item can be described with a *mixture distribution*; if T is the product lifetime, then its lifetime distribution $F(t)$ is extricated to

$$F(t) = \omega F_a(t) + (1 - \omega)F_0(t), \tag{33}$$

where F_0 is the lifetime distribution of the normal (non-defective) products, ω is the proportion of defective (or non-conforming) products that have distribution F_a, where $F_a(t) > F_0(t)$. Manufacturers of large, repairable systems, including the automobile industry, can benefit greatly by quickly identifying a finished product that was generated from the nonconforming population F_a and getting it out of service as soon as possible.

In general, the defective items are costly to the manufacturer thereby greatly influencing the warranty policy and limiting the protection the manufacturer will offer to the consumer. Mixtures have been helpful in modeling repair times for warranty policy, including heuristic models by Majeske and Herrin (1995). Majeske (2003) used a mixture hazard function to model the time to first warranty claim, and estimated the fraction of vehicles containing a manufacturing or assembly defect when leaving the assembly plant.

The repair process is modeled as a *minimal repair process* generated from the mixture in (33). Once the system fails, it is automatically repaired to be as good as an identical system that has survived to the same age. The resulting sequence of failure times constitutes a nonhomogeneous Poisson process with mean rate function equal to the underlying cumulative hazard rate. Obviously, if the system has a greatly increasing rate of failure, the overall cost of operating the system is strongly dependent on the replacement policy. Kvam, Singh, and Whitaker (2002) considered estimating the system lifetime distribution in the case the system was known to have an increasing failure rate.

For practical consideration, we focus on the non-homogeneous Poisson process with intensity function

$$v(t) = \frac{\beta}{\theta}(t/\theta)^{\beta-1} \tag{34}$$

which is commonly accepted as an effective model for many repairable systems, e.g., see Rigdon and Basu (2000). A convenient alternative parametrization for (34) is

$$v(t) = \lambda\beta t^{\beta-1}. \tag{35}$$

This model is called the *power-law process* (PLP) because the intensity function is proportional to a power of t. We call λ the intensity parameter, β the shape parameter, and θ the scale parameter. The Power-Law process is frequently used to model repairable system lifetimes, as evident in in Duane (1964), Ridgon, et al. (1998), and Ridgon and Basu (2000). Engelhardt & Bain (1987) used a compound power-law model to characterize the heterogeneity of different systems by treating λ as a random variable from the gamma distribution. This frailty-type model accounts for general heterogeneity of the population, but is not effective in modeling nonconforming systems. In this chapter, we choose to model multiple systems as mixture power-law processes with two point mixture distributions. These correspond to two types of intensity functions, $v_0(t)$ and $v_a(t)$ for conforming and nonconforming systems, respectively. The higher failure rate of the nonconforming subpopulation is characterized by an inequality between their respective intensity parameters: $\lambda_a > \lambda_0$.

Consider n manufactured systems with intensity function $v_i(t) = \lambda_i \beta_i t^{\beta_i - 1}$, $i = 1, ..., n$. The systems are possibly time truncated or failure truncated. For time truncated systems, we observe system i over time interval $(0, \tau_i)$; τ_i may be the current calendar time. Denote t_{ij} as the j^{th} failure time for system i, and $j = 1, \ldots, k_i$, where k_i is the number of failures before censoring time τ_i.

For the failure truncated case, a system is taken off test after a fixed number of failures is observed. Denote k_i as the pre-fixed number of failures, then the failure times t_{ij}'s are recorded for $j = 1, \ldots, k_i$. In the example that follows, the data can be time truncated or failure truncated. The detection of a PLP is shown in Section 2 by using copy machine failure times as an example. The copy machines exhibit a PLP mixture of two intensity parameters (and a single shape parameter β). Estimation, based on maximum likelihood, is described in Section 3. In Section 4, we use these estimates to develop an optimal strategy for warranty decision making.

4.2 Exploratory Study of Copy Machine Failure

Figure 20 shows the failure-time data for a group of 20 copy machines (Zaino and Berke, 1992). For these machines, time is measured by the number of *actuations*, i.e., the number

of copies made, and the time at installation is defined to be 0. This data set (adjusted for staggered installation times) is displayed in Table 4. Copiers removed from the test upon 8 failures were failure truncated, while other copiers are regarded as time censored at $\tau = 40000$ actuations.

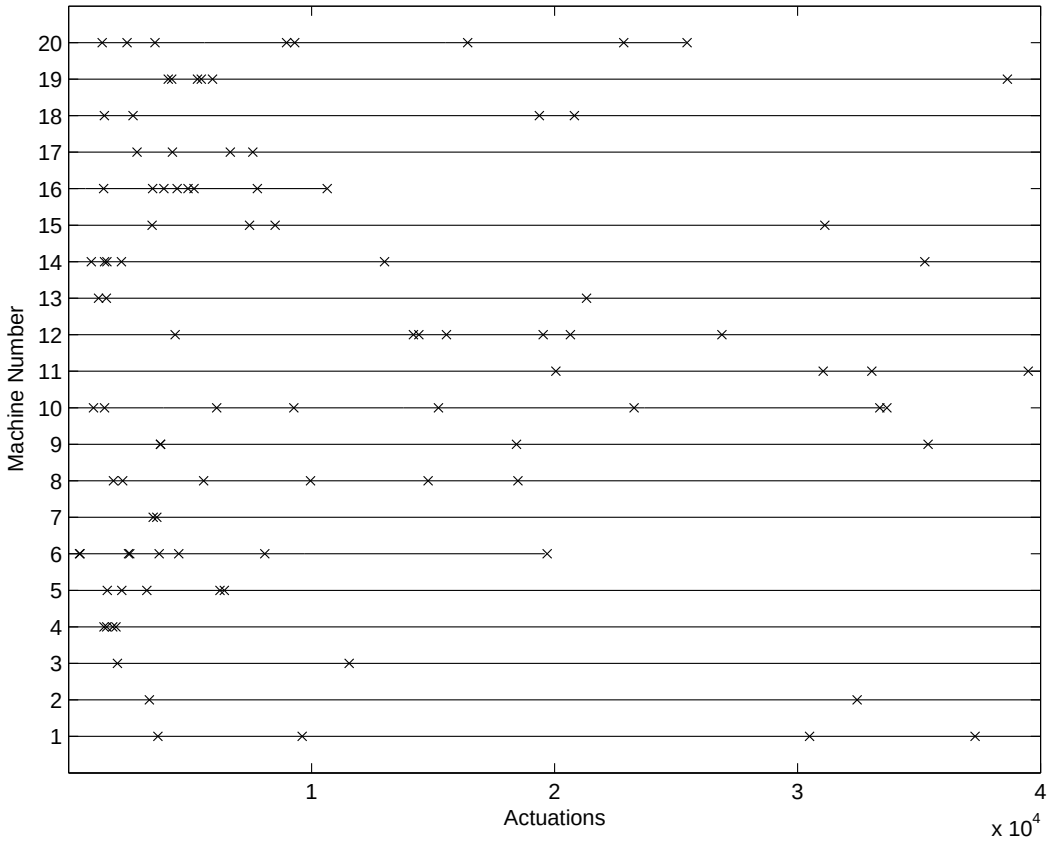

Figure 20: Number of actuations between failures for 20 tested copy machines. Data from Zaino and Berke (1992).

For failure time T_i and number of failures K_i, we use the notation of Rigdon et al. (1998) for cases where some systems are failure truncated and others are time truncated:

$$T_i = \begin{cases} \tau_i & \text{if system } i \text{ is time truncated} \\ t_{i,k_i} & \text{if system } i \text{ is failure truncated} \end{cases}$$

$$K_i = \begin{cases} k_i & \text{if system } i \text{ is time truncated} \\ k_i - 1 & \text{if system } i \text{ is failure truncated.} \end{cases}$$

Table 4: Number of actuations until failure for copy machine failure data

i	t_{i1}	t_{i2}	t_{i3}	t_{i4}	t_{i5}	t_{i6}	t_{i7}	t_{i8}
1	3678	9619	30497	37308	-	-	-	-
2	3328	32456	-	-	-	-	-	-
3	2016	11551	-	-	-	-	-	-
4	1463	1570	1820	1956	-	-	-	-
5	1596	2189	3219	6233	6409	-	-	-
6	452	472	2467	2517	3727	4537	8079	19694
7	3487	3635	-	-	-	-	-	-
8	1847	2230	5557	9958	14795	18494	-	-
9	3783	3787	18436	35375	-	-	-	-
10	1027	1483	6101	9269	15225	23273	33389	33675
11	20057	31058	33061	39497	-	-	-	-
12	4390	14190	14420	15550	19535	20650	26890	-
13	1233	1555	21318	-	-	-	-	-
14	940	1479	1583	2177	13004	35241	-	-
15	3439	7451	8503	31126	-	-	-	-
16	1443	3464	3926	4473	4918	5161	7768	10649
17	2818	4276	6656	7581	-	-	-	-
18	1474	2653	19378	20816	-	-	-	-
19	4105	4247	5305	5466	5924	38635	-	-
20	1382	2409	3557	8974	9312	16429	22850	25455

Based on the likelihood for an individual system,

$$L(\lambda_i, \beta_i) \propto \exp(-\lambda_i T_i^{\beta}) \prod_{j=1}^{k_i} \lambda_i \beta_i t_{ij}^{\beta_i - 1}, \tag{36}$$

the maximum likelihood estimators (MLEs) $\hat{\lambda}_i$ and $\hat{\beta}_i$ can be obtained as

$$\hat{\beta}_i = \frac{k_i}{\sum_{j=1}^{k_i} \log(T_i/t_{ij})}, \ \hat{\lambda}_i = \frac{T_i}{k_i^{1/\hat{\beta}_i}}. \tag{37}$$

To obtain a more parsimonious model, we test equality of the intensity functions for individual systems. The shape parameter β demonstrates the reliability development efforts, i.e., $\beta > 1$ shows system reliability decreasing in time and $\beta < 1$ shows reliability growth. With the MLE $\hat{\beta}_i$ from (37), it is well known (Chapter 4 of Rigdon & Basu, 2000) that the conditional distributions of the variables $2k_i\beta_i/\hat{\beta}_i$, $i = 1, \ldots, n$ given $k_1, \ldots, k_n$ are independent and chi-squared with $2K_i$ degrees of freedom. The $100(1-\alpha)\%$ confidence intervals for β_i's are given as

$$\left(\frac{\chi^2_{\alpha/2}(2K_i)\hat{\beta}_i}{2k_i}, \frac{\chi^2_{1-\alpha/2}(2K_i)\hat{\beta}_i}{2k_i} \right),$$

where $\chi^2_{1-\alpha/2}(2K_i)$, and $\chi^2_{\alpha/2}(2K_i)$ are the $1-\alpha/2$ and $\alpha/2$ quantiles for chi-square distribution with $2K_i$ degrees of freedom.

The hypothesis $\beta_i = \beta$ implies that reliability development efforts are equally effective for systems being tested. Crow (1974) suggests a likelihood ratio test for testing the equality of β's,

$$H_0 : \ \beta_1 = \beta_2 = \cdots = \beta_n,$$

against the alternative that at least two of the β's are different based on

$$\mathbf{LR} = \sum_i \beta^* - \sum_i k_i \log \hat{\beta}_i,$$

where $\hat{\beta}_i$ is the MLE for β_i and β^* is the weighted mean of the $\hat{\beta}_1, ..., \hat{\beta}_n$:

$$\beta^* = \frac{\sum_{i=1}^n k_i}{\sum_{i=1}^n k_i/\hat{\beta}_i}.$$

Using an approximation similar to Bartlett's (1937) statistic testing for equal variances in independent normal distributions, the null distribution for the test statistic $-2 \times \mathbf{LR}/a$ is

$\chi^2(n-1)$, where $a = 1 + \left(\sum_{i=1}^n 1/k_i - 1/(\sum_{i=1}^n k_i)\right)/(6(k-1))$. This test is applied for the copy machine failure data in Table 4, with p-value = 0.59; there is no strong evidence for modeling the shape parameters differently.

Given the shape parameter β is identical for all systems, we can proceed to test the equality of λ_i's. Under $H_0 : \lambda_i = \lambda$, the likelihood function is

$$L(\lambda, \beta) \propto \prod_{i=1}^{n} \left\{ \exp(-\lambda_i T_i^{\beta}) \prod_{j=1}^{k_i} \lambda_i \beta t_{ij}^{\beta-1} \right\}.$$

The MLEs for λ_i and β satisfy the following estimating equations:

$$\begin{aligned} \lambda_i &= \frac{\sum_{i=1}^n k_i}{\sum_{i=1}^n T_i^{\beta}} \\ \frac{\sum_{i=1}^n k_i}{\beta} &= \left\{ \sum_{i=1}^{n} \lambda_i T_i^{\beta} \log(T_i) - \sum_{i=1}^{n} \sum_{j=1}^{k_i} \log(t_{ij}) \right\}. \end{aligned}$$

If all the n systems are time truncated at τ, then β is solved explicitly as

$$\hat{\beta} = \frac{\sum_{i=1}^{n} k_i}{\sum_{i=1}^{n} \sum_{j=1}^{k_i} \log(\frac{\tau}{t_{ij}})}.$$

In other cases, explicit solutions for β and λ_i are not guaranteed.

Lee (1980) proposed a test for comparing rates of several independent PLP processes. A test can be constructed based on the count data k_i when β_i are assumed to be the same. Conditional on the total number of failure times $k = \sum_{i=1}^n k_i$, the distribution of the failures counts K_i is multinomial with cell probabilities

$$\pi_i = \frac{\psi_i \tau_i^{\beta}}{\sum_{i=1}^n \psi_i \tau_i^{\beta}} \tag{38}$$

and the problem is reduced to testing multinomial parameters with $H_0 : \pi_1 = \pi_2 = ... = \pi_n$ (versus H_a : some π_i are not equal) on the simplex $\sum \pi_i = 1$. Let β_n be a consistent estimator of β (this is explained in the next section). A test for homogeneity based on (38) can be constructed from

$$\hat{\pi}_i = \frac{\tau_i^{\beta_n}}{\sum_{i=1}^n \tau_i^{\beta_n}},$$

with corresponding dispersion statistic

$$q_n = \sum_{i=1}^{n} \frac{(k_i - k\hat{\pi}_i)^2}{k\hat{\pi}_i}.$$

Under H_0, q_n has a limiting χ^2 distribution with $n-1$ degrees of freedom. For the copy machine data, the test statistic q_n is calculated to be 33.84, and the hypothesis of homogeneity for λ is rejected with a p-value 0.019.

A graphical plot can be applied to detect the heterogeneity in the intensity parameters if the number of failures is large enough. Conditional on β and λ_i, the total number of failures for i^{th} system, K_i, is a Poisson random variable with mean $\lambda_i T_i^{\beta}$. Then, if K_i is sufficiently large, the transformed count data

$$Z_i = \frac{K_i - \lambda T_i^{\beta}}{\sqrt{\lambda T_i^{\beta}}}, \tag{39}$$

is approximately distributed as standard normal distribution under H_0 where $\lambda_i = \lambda$. Hence, after replacing λ and β by their consistent estimators, a normal plot for Z_i can be used to examine the homogeneity for λ_i. To show how this procedure works, we use a simple PLP simulation below.

Simulation Example 1: We simulate a mixture Power Law process with $n = 200$ systems, using the simulation procedure given in Meeker & Escobar (1998, page. 418). The proportion for nonconforming systems ω is set to be 0.05; the intensity parameters for conforming and nonconforming systems are $\lambda_0 = 1$, $\lambda_a = 5$, respectively. The common shape parameter β is chosen to be 1.5, indicating reliability deterioration and the censoring time $\tau_i = \tau = 4$, is the same for all the systems.

Figure 21(a) shows the normal plots of the transformed Z_i's for the mixture population in the simulation, where a lack-of-fit can be detected visually. The normal plot for the conforming systems is shown in Figure 21(b), which has no strong visual evidence for lack-of-fit.

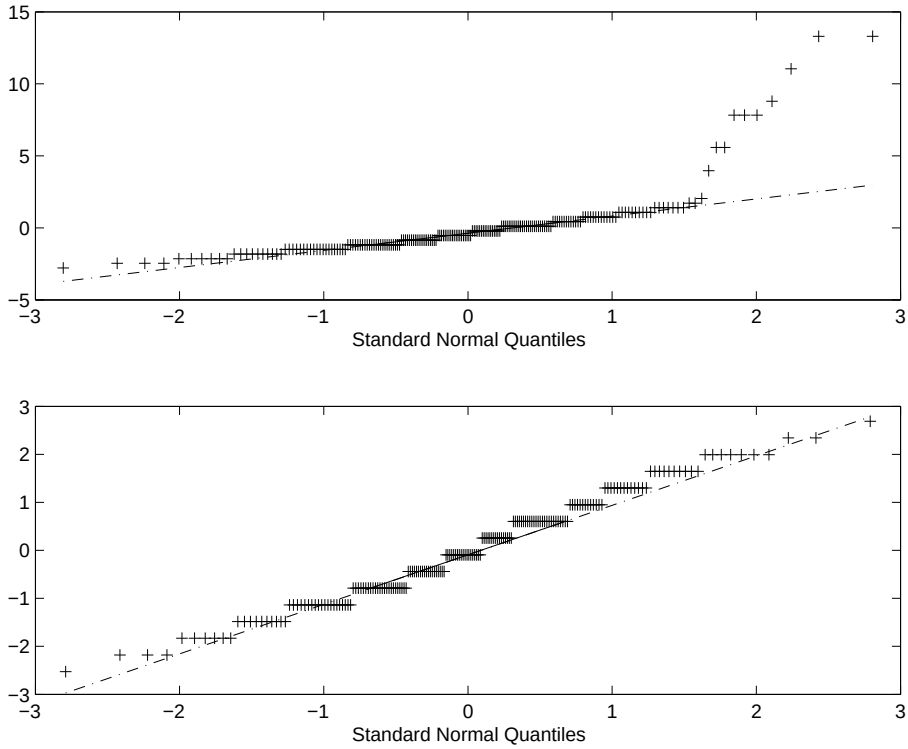

Figure 21: The Normal QQ Plot for the transformed counts data Z in Simulation Example 1

4.3 Mixture Model

After testing the copy machine failure data for goodness-of-fit, we assume the regular population is related to the non-conforming population via a common shape parameter for the joint processes modeled with intensity functions $v_0(t) = \lambda_0 \beta t^{\beta-1}$ and $v_a(t) = \lambda_a \beta t^{\beta-1}$. We next consider the mixture model to describe these two sub-populations.

4.3.1 PLP likelihood for mixture

The likelihood based on the failure data $\{t_{ij}, 1 \leq i \leq m \text{ and } 1 \leq j \leq k_i\}$ is a function of the parameters of the PLP intensity function, $\{\lambda_0, \lambda_a, \beta\}$. That is, the shape parameter β is the same for both $v_0(t)$ and $v_a(t)$. The mixing parameter ω is the proportion of non-conforming items in the population, and is assumed to be small ($\omega \ll 0.5$). The intensity functions for conforming and nonconforming systems are $v_0(t) = \lambda_0 \beta t^{\beta-1}$ and $v_a(t) = \lambda_a \beta t^{\beta-1}$, respectively. Then the likelihood function is

$$L(\boldsymbol{\theta};\boldsymbol{t}) \propto \prod_{i=1}^{n} \left\{ (1-\omega)\lambda_0^{k_i}\beta^{k_i}\exp(-\lambda_0\tau_i^{\beta}) \prod_{j=1}^{k_i} t_{ij}^{\beta-1} + \omega\lambda_a^{k_i}\beta^{k_i}\exp(-\lambda_a\tau_i^{\beta}) \prod_{j=1}^{k_i} t_{ij}^{\beta-1} \right\} \quad (40)$$

where $\boldsymbol{t} = \{t_{ij}\},\ i = 1, \ldots, n\ j = 1, \ldots, k_i$, and $\boldsymbol{\theta} = \{\lambda_0, \lambda_a, \omega, \beta\}$.

Obviously, there is no general closed from solution in (40) for the MLE of $\boldsymbol{\theta}$. To set up a simple iterative method for solving the MLE, the EM-Algorithm (see McLachlan and Krishnan (1996), for example) can be applied by defining the unobserved quantity z_i, where $z_i = 0$ if the i^{th} system is from the conforming population ($z_i = 1$ otherwise), so that $P(Z_i = 1) = \omega$.

With $\boldsymbol{z} = \{z_i,\ i = 1, ..., m\}$, the "full data" likelihood (including $\boldsymbol{z}$) is relatively simple and well behaved:

$$L(\boldsymbol{\theta};\boldsymbol{t},\boldsymbol{z}) \propto \prod_{i=1}^{n} \left\{ \lambda_0^{k_i}\beta^{k_i}\exp(-\lambda_0\tau_i^{\beta}) \prod_{j=1}^{k_i} t_{ij}^{\beta-1} \right\}^{1-z_i} \left\{ \lambda_a^{k_i}\beta^{k_i}\exp(-\lambda_a\tau_i^{\beta}) \prod_{j=1}^{k_i} t_{ij}^{\beta-1} \right\}^{z_i}. \quad (41)$$

The EM algorithm solves for the MLE by estimating $\boldsymbol{z}$ (or a function of $\boldsymbol{z}$ determined through the log-likelihood) and maximizing over the simpler likelihood in (41) by treating the estimated values of $\boldsymbol{z}$ as observed data. The algorithm consists of two steps: the *E-step* (estimating $\boldsymbol{z}$) and the *M-step* (finding the MLE using the estimates in the E-step).

E-step: In the p^{th} iteration, z_i is replaced (estimated) by its expected value $\xi_i^{(p)}$ in the full likelihood (41), given current parameter estimates $\lambda_0^{(p)}, \lambda_a^{(p)}, \omega^{(p)}, \beta^{(p)}$ where

$$P(Z_i = r) = \begin{cases} \omega^{(p)}\exp(-\lambda_a^{(p)}\tau_i^{\beta^{(p)}}) \prod_{j=1}^{k_i} v_a^{(p)}(t_{ij}) & \text{if } r = 1 \\ (1-\omega^{(p)})\exp(-\lambda_0\tau_i^{\beta^{(p)}}) \prod_{j=1}^{k_i} v_0^{(p)}(t_{ij}) & \text{if } r = 0 \end{cases} \quad (42)$$

M-step: By setting the first derivative of the full log-likelihood function from (41) to zero,

we generate the following estimating equations:

$$\lambda_a^{(p+1)} = \frac{\sum_{i=1}^{n} \xi_i^{(p)} k_i}{\sum_{i=1}^{n} \xi_i^{(p)} \tau_i^{\beta^{(p+1)}}}, \quad \lambda_0^{(p+1)} = \frac{\sum_{i=1}^{n} (1-\xi_i^{(p)}) k_i}{\sum_{i=1}^{n} (1-\xi_i^{(p)}) \tau_i^{\beta^{(p+1)}}}$$

$$\frac{\sum_{i=1}^{n} k_i}{\beta^{(p+1)}} = \left\{ \sum_{i=1}^{n} (1-\xi_i^{(p)}) \lambda_0^{(p)} \tau_i^{\beta^{(p+1)}} \log(\tau_i) + \xi_i^{(p)} \lambda_a^{(p)} \tau_i^{\beta^{(p+1)}} \log(\tau_i) \right\} - \sum_{i=1}^{n} \sum_{j=1}^{k_i} \log(t_{ij}) \quad (43)$$

and ω is updated as $\omega^{(p+1)} = \sum_{i=1}^{n} \xi_i^{(p)}/M$. The E-step and the M-step are repeated until the parameter estimates converge to the MLEs. In this case, convergence is guaranteed by Theorem 2 in Wu (1983) because the full-data likelihood is a member of the exponential family.

For the copier data, the EM steps were repeated until the parameter estimates converged to stationary points, which can be monitored by the trace of the algorithm output. The MLEs are $(\hat{\lambda}_0, \hat{\lambda}_a, \hat{\beta}, \hat{\omega}) = (0.0091, 0.0229, 0.5862, 0.1439)$. The result shows that the systems are experiencing reliability growth by the fact $\hat{\beta} = 0.58 < 1$; about 14.4% of the total population seems to come from a subpopulation with higher failure rate. The ξ_i's from the EM Algorithm can be regarded as the posterior probability of being in the nonconforming group for system i. Based on a simple rule by classifying a system as nonconforming if $\xi_i > 0.5$ (this would obviously change if a non-degenerate risk function were used), machines 6, 16, 20 are classified as nonconforming by the fact that $\xi_6 = 0.74$, $\xi_{16} = 0.9174$, and $\xi_{20} = 0.5760$.

4.3.2 PLP Model Inference

Titterington (1990) has shown that inference for mixture distributions can be fraught with problems of non-identifiability and unsolvable likelihoods. In this case, we are assuming the mixture has two components, which greatly simplifies the problem structure. For testing $H_0 : \omega = 0$ versus $H_a : \omega > 0$, the likelihood ratio

$$\Lambda = \frac{\sup_{H_0} L(\boldsymbol{\theta}; \boldsymbol{t})}{\sup_{H_a} L(\boldsymbol{\theta}; \boldsymbol{t})} \quad (44)$$

is simple enough to compute. Under standard regularity conditions for the likelihood (see Lehmann, 1997, for example), $X^2 = -2\log\Lambda$ is distributed ${\chi_1}^2$. However, likelihood based procedures are not guaranteed even in this case; the regularity conditions on the parameter space that satisfy requirements for MLE limit properties cannot be met. For the null hypothesis of homogeneity, the parameter space includes parameter boundary values $\omega = 0$ along with the line $\lambda_0 = \lambda_a$, corresponding to a non-identifiable subset of the parameter space $\Theta = \{(\omega, \lambda_0, \lambda_a, \beta) \in ([0,1], (\Re^+)^3)\}$.

In place of a conventional likelihood ratio test, computational methods can be used for tests and confidence regions for unknown parameters based on resampling methods as demonstrated in Feng & McCulloch (1996). For the hypotheses

$$H_0 : v(t) = v_0(t) \quad \text{versus} \quad H_a : v(t) = (1-\omega)v_0(t) + \omega v_a(t),$$

an approximate test is constructed by the following bootstrap likelihood ratio procedure:

1. Compute the MLE $\hat{\boldsymbol{\theta}}_0$ of $\boldsymbol{\theta}_0 = (\lambda, \beta)$ under H_0.

2. Generate a bootstrap sample corresponding to the $\hat{v}_0(t)$, where the unknown parameters are replaced by the MLE $\hat{\boldsymbol{\theta}}_0$.

3. Compute the test statistic $X^2 = -2\log\Lambda$ corresponding to (44) after finding two sets of MLEs.

4. Repeat these last two steps B times ($B > 1000$, at least) and store the B values of the test statistics ${X_1}^2, ..., {X_B}^2$.

5. Compute the significance of X^2 using the distribution of the B test statistics as the null distribution.

¿From these steps, the replicated values of $-2\log\Lambda$ formed from the successive bootstrap samples provide an assessment of the bootstrap, i.e., the null distribution of $-2\log\Lambda$. The j^{th} order statistic in the B replications can be taken as an estimate of the $100j/B$ percentile of the null distribution. Thus, the p-value can be approximated by comparing the bootstrapped samples with the original X^2 test statistic.

The bootstrap approach can also be used to study the standard errors of the MLE for $\boldsymbol{\theta} = (\omega, \lambda_a, \lambda_0, \beta)$. A simple nonparametric bootstrap is applied here to avoid the complexity of simulating the nonhomogeneous Poisson process. We first construct B bootstrap samples $\boldsymbol{t}_1^*, \boldsymbol{t}_2^*, \ldots, \boldsymbol{t}_B^*$ by resampling with replacement from the n observation systems. Let $\hat{\boldsymbol{\theta}}_1^*, \ldots, \hat{\boldsymbol{\theta}}_B^*$ be the bootstrap estimates of $\boldsymbol{\theta}$ calculated from $\boldsymbol{t}_1^*, \ldots, \boldsymbol{t}_B^*$, respectively, using the EM algorithm. The covariance matrix of $\hat{\boldsymbol{\theta}}$ can be estimated using the sample covariance matrix of $\hat{\boldsymbol{\theta}}_1^*, \ldots, \hat{\boldsymbol{\theta}}_B^*$,

$$V = \sum_{k=1}^{B} (\hat{\boldsymbol{\theta}}_k^* - \bar{\boldsymbol{\theta}}^*)(\hat{\boldsymbol{\theta}}_k^* - \bar{\boldsymbol{\theta}}^*)^T/(B-1),$$

where $\bar{\boldsymbol{\theta}}^* = \sum_{k=1}^{B} \hat{\boldsymbol{\theta}}_k^*/B$.

Under H_0, the repair data for copy machine failures lead to $(\hat{\lambda}, \hat{\beta}) = (0.0134, 0.5639)$, and the log likelihood ratio is calculated as $X^2 = 2.4756$. Based on $B = 2000$ bootstrap samples representing the null distribution, the p-value for the original repair data is 0.32. This lack of strong evidence is due, in part, to the small sample size of $n = 20$ for the mixture problem.

For Simulation Example 1, the histogram for model parameters using nonparametric bootstrap method is shown in Figure 22. The histograms show that all the distributions are approximately symmetric.

Remark 1: The exact point estimates for both parameters as well as an exact interval estimate for the shape parameter for a single system are well studied (Finkelstein, 1976). For multiple systems with identical λ and β, the asymptotic properties for MLEs $\hat{\lambda}$ and $\hat{\beta}$ can be derived. To keep this presentation short, we consider the case where all the systems are failure truncated on the right with the same failure number m. By letting the number of systems $n \to \infty$ and the failure number $m \to \infty$, the asymptotic confidence intervals for $\hat{\lambda}$ and $\hat{\beta}$ can be obtained from the Fisher information matrix as shown in Theorem 1 below.

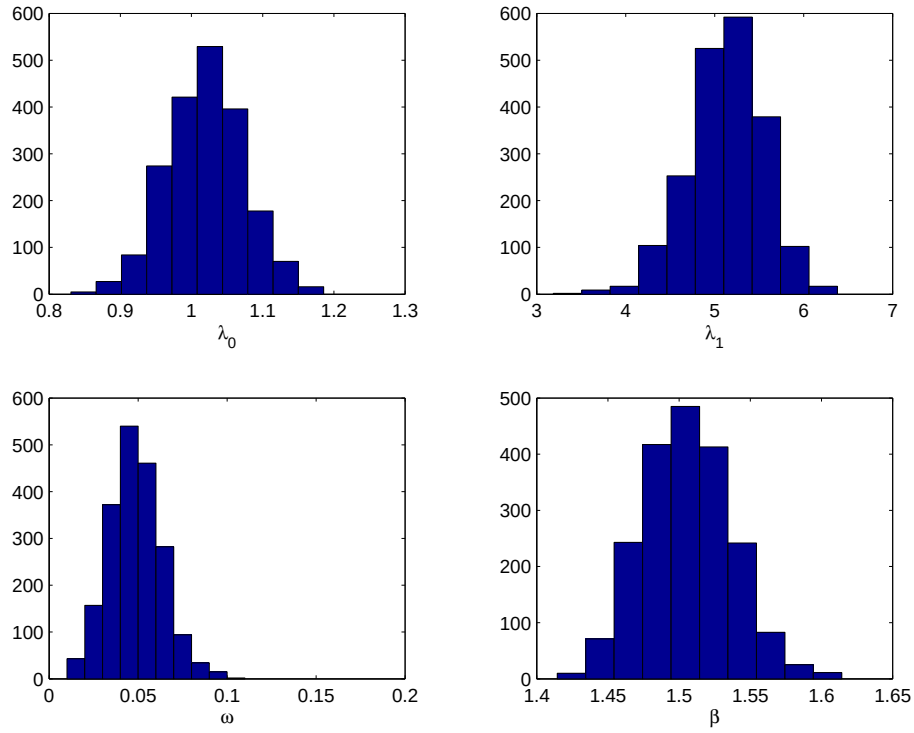

Figure 22: The Histograms for the model parameters in mixture Power Law Processes based on bootstrapped samples in Simulation Example 1.

Remark 2: For the homogeneous population, another estimator for the shape parameter

$$\tilde{\beta} = \frac{\sum_{i=1}^{n} k_i}{\sum_{i=1}^{n}\sum_{j=1}^{k_i} \log(\frac{T_i}{t_{ij}})} \tag{45}$$

is called the *conditional* MLE; Rigdon, et al. (1998) showed that conditional on system i having K_i failures, the random variable $2K\beta/\tilde{\beta}$ has an approximate χ^2 distribution with $2K$ degrees of freedom where $K = \sum_{j=1}^{n} K_i$. The transformed random variables $U_{ij} = \log(T_i/T_{i,k_i-j+1})$ are distributed as K_i order statistics from an exponential distribution with (unknown) mean $1/\beta$. The standard estimator for the mean of U_{ij} is $\sum_{i=1}^{m}\sum_{j=1}^{K_i} U_{ij} / \sum_{i=1}^{m} K_i$, which simplifies to $1/\tilde{\beta}$.

By extending the limit results of Gaudoin, et al. (2004) to multiple systems, the asymptotic normality of the MLEs can be obtained, allowing hypothesis tests and confidence regions to be constructed via the Fisher information. The proof of the theorem that follows is relegated to the Appendix.

Theorem 1. *For i independent systems, let t_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, m$ be the failure times from system i, where failure times are governed by a Power Law Process with parameter vector $\boldsymbol{\theta} = (\lambda, \beta)$. Then, under the standard regularity conditions for MLEs, as $n \to \infty$ and $m \to \infty$,*

$$\sqrt{nm}(\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}) \to N(0, \mathcal{I}(\bar{\boldsymbol{\theta}})^{-1}),$$

where

$$\mathcal{I}(\bar{\boldsymbol{\theta}})^{-1} = \begin{pmatrix} \lambda^2[1 + (\log \frac{m}{\lambda})]^2 & -\lambda\beta \log \frac{m}{\lambda} \\ -\lambda\beta \log \frac{m}{\lambda} & \beta^2 \end{pmatrix} \tag{46}$$

is the inverse of the Fisher information matrix.

4.4 Optimal Strategy in Warranty Decision Making

Suppose that from the recent repair history of a group of similar systems, we know the intensity parameters for the nonconforming and conforming systems are λ_a and λ_0, respectively. Further suppose that under the minimal repair warranty policy, failed products experience minimal repair without any cost to the consumers, but the manufacturer incurs a cost of $C_m > 0$ per repair. Let t_w be the length of the warranty coverage. Then the expected total repair costs for conforming systems and nonconforming systems are $C_m\Lambda_0(t_w)$ and $C_m\Lambda_a(t_w)$, respectively, where $\Lambda_0(t_w) = \lambda_0 t_w^\beta$, and $\Lambda_a(t_w) = \lambda_a t_w^\beta$. If the minimal repair costs for nonconforming products are high enough (compared to the fixed cost C_T of system replacement), we can lower the total repair costs by identifying and removing those nonconforming systems before t_w.

Consider the case where the products are examined after k failures, i.e., the product lifetimes are failure truncated on the right. We classify the products into two groups based on the hypothesis test $H_0 : \lambda_i = \lambda_0$ vs. $H_a : \lambda_i = \lambda_a$. The expected costs due to the classification errors are given in Table 5. Denote $P(H_a|H_0)$ and $P(H_0|H_a)$ as the Type I and Type II errors, respectively. The total expected cost function is:

$$\begin{aligned} C(k) &= m \times (1 - \omega) \times P_k(H_a|H_0) \left\{ (C_T + C_m k) - C_m \lambda_0 t_w^\beta \right\} \\ &+ m \times \omega \times P_k(H_0|H_a) \left\{ C_m \lambda_a t_w^\beta - (C_T + C_m k) \right\}, \end{aligned} \tag{47}$$

where $0 \leq k \leq \lambda_a t_w^\beta - C_T/C_m$ and $\lambda_0 t_w^\beta < C_T/C_m$, since the misclassification costs will always be larger than 0.

Table 5: The Cost Functions for Misclassifications

prob	cost function
$P(H_0\|H_0)$	$C_m\lambda_0 t_w^\beta$
$P(H_a\|H_0)$	$C_m k + C_T$
$P(H_0\|H_a)$	$C_m\lambda_a t_w^\beta$
$P(H_a\|H_a)$	$C_m k + C_T$

Corresponding to the hypothesis test $H_0 : \lambda_i = \lambda_0$ versus $H_a : \lambda_i = \lambda_a$, the likelihood ratio statistic is

$$\begin{aligned} \mathbf{LR} &= \frac{\exp(-\lambda_a t_{ik}^\beta)\prod_{j=1}^k \lambda_a \beta t_{ij}^{\beta-1}}{\exp(-\lambda_0 t_{ik}^\beta)\prod_{j=1}^k \lambda_0 \beta t_{ij}^{\beta-1}} \\ &= \left(\frac{\lambda_a}{\lambda_0}\right)^k \exp[(\lambda_0 - \lambda_a)t_{ik}^\beta]. \end{aligned}$$

The uniformly most powerful (UMP) test (Lehmann, 1997, page 74) is to reject H_0 if $t_{ik} < \eta_k$, where η_k is the critical value to be decided. Under H_0, t_{ik} has a generalized gamma distribution GGAM(λ, β, k) (see Ridgon & Basu, 2000, page 57) with cumulative distribution function G given as

$$G(t; \lambda, \beta, k) = \Gamma_I(\lambda t^\beta; k),$$

and Γ_I is the incomplete gamma function defined by $\Gamma_I(v; k) = \int_0^v x^{k-1}\exp(-x)dx/\Gamma(k)$, $v > 0$. By controlling the Type I error level at α, the critical value η_k can be solved from

$$\Gamma_I(\lambda_0 \eta_k^\beta; k) = \alpha.$$

Then the Type II Error can be calculated as

$$\begin{aligned} P_k(H_0|H_a) &= 1 - \Gamma_I(\lambda_a \eta_k^\beta; k) \\ &= 1 - \Gamma_I[\Gamma_I^{-1}(\alpha; k)\lambda_a/\lambda_0], \end{aligned} \tag{48}$$

where $\Gamma_I^{-1}(\,\cdot\,; k)$ is the inverse function of $\Gamma_I(\,\cdot\,; k)$. By plugging in $P_k(H_0|H_a)$ from (48)into (47), we can minimize the expected cost numerically with $k \in [1, \lambda_a t_w^\beta - C_T/C_m]$.

Observation 1: If ω is small such that the nonconforming products do not affect the total costs (47) as much as conforming products, $C(k)$ is an increasing function in k, and the manufacturer benefits from earlier testing.

Observation 2: Figure 23 shows the Type II error as function of the ratio λ_a/λ_0 under $\alpha = 0.05$. If the ratio $\lambda_a/\lambda_0 > 5$, we can see that the Type II error approaches 0 quickly as k increases. When λ_a/λ_0 is large, nonconforming products are more easily detected even without a large failure number k.

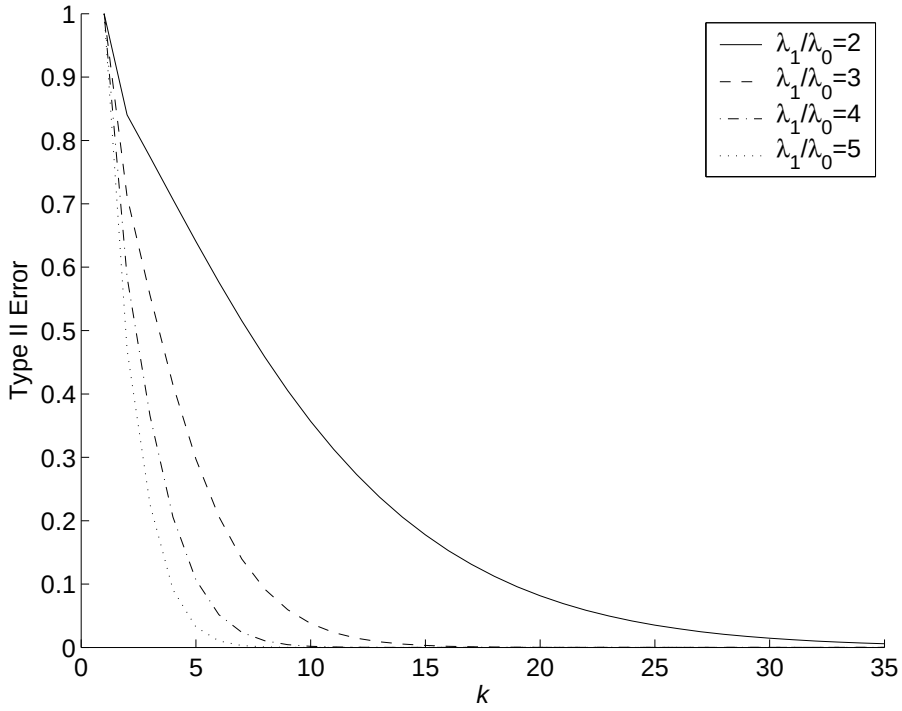

Figure 23: Type II Error function for Hypothesis Testing using different ratios of the intensity parameters.

Simulation Example 2: We illustrate this optimal decision process through the following simulation. The total warranty coverage time set as $t_w = 4$ years, and we use four different values of β (i.e., $\beta \in \{0.5, 1, 1.5, 2\}$) to illustrate reliability varying from reliability growth to deterioration.

The proportion of nonconforming systems ω is set to one of four different values ($\omega \in \{0.001, 0.01, 0.05, 0.1\}$) to compare different scenarios of population contamination. Without

loss of generality, λ_0 is chosen to be 2, and corresponding to this, λ_a is set to be three, five or ten times the value of λ_0, i.e., $\lambda_a \in \{3\lambda_0, 5\lambda_0, 10\lambda_0\}$). Finally, C_T/C_m is assumed to be equal to $(\lambda_a + \lambda_0)t_w^\beta/2$. The optimal values of k under each of the different cases are shown in Table 6, which supports Observations 1 and 2 above. For example, with $\lambda_a = 3\lambda_0$, $\beta = 2$ and $\omega = 0.05$, the expected cost is minimized by choosing $k = 8$. That is, the optimal strategy is to test each system until eight failures occur before deciding whether or not the system is conforming or nonconforming.

Table 6: The Optimal k under different model parameters in simulated process

$\lambda_a/\lambda_0 = 3$				
ω	$\beta = 0.5$	$\beta = 1$	$\beta = 1.5$	$\beta = 2$
0.001	1	1	1	1
0.01	1	1	1	3
0.05	2	4	6	8
0.1	4	6	8	10
$\lambda_a/\lambda_0 = 5$				
ω	$\beta = 0.5$	$\beta = 1$	$\beta = 1.5$	$\beta = 2$
0.001	1	1	1	1
0.01	1	2	3	4
0.05	3	4	5	6
0.1	4	5	6	6
$\lambda_a/\lambda_0 = 10$				
ω	$\beta = 0.5$	$\beta = 1$	$\beta = 1.5$	$\beta = 2$
0.001	1	1	1	2
0.01	2	2	3	3
0.05	3	3	3	4
0.1	3	3	4	4

4.5 Conclusion

This chapter studies the modeling of heterogenous systems governed by a minimal repair process. An exploratory study and graphical methods are used to detect heterogeneity of the power law processes for 20 copy machines based on repeated failure-time data. Bootstrap methods are used to calibrate the estimation uncertainty as well as likelihood ratio test statistics.

When considering a model for conforming and noncomforming systems, the two-point

mixture model makes intuitive sense and is easily interpreted. Furthermore, it lends itself to a natural formula for classifying products as non-conforming or conforming. However, discrete mixtures are difficult to fit, especially with small samples. Alternatively, the *continuous* mixture model generated with a Gamma mixing distribution for λ (Englehardt and Bain, 1987) will fit the copy machine failure data, but the estimated mixing parameters from the Gamma distribution are poorly fit, especially the shape parameter. This is due, in part, to the small sample size.

Finally, an optimal decision based on estimated values is derived to minimize warranty cost. The decision process is aided by "missing data" estimates in the EM Algorithm. Future study can consider more complex warranties based on intricate risk functions. Our asymptotic results are based on a simple system of minimal repair with failure truncation on the right, and confidence statements for the power law process parameters can be constructed from the Fisher Information matrix of Theorem 1.

4.6 Appendix: Proof of Theorem 1

The asymptotic normality of the parameter estimates for a single system is demonstrated in Gaudoin, et al. (2004). To shorten this presentation, we only illustrate the derivation of the asymptotic covariance through the information matrix. The likelihood for the repair times can be expressed as

$$L(\beta, \lambda; \boldsymbol{t}) \propto \prod_{i=1}^{n} \left\{ \exp(-\lambda t_{im}^{\beta}) \prod_{j=1}^{m} \lambda \beta t_{ij}^{\beta-1} \right\}$$

and the corresponding Fisher Information matrix is obtained as:

$$\mathcal{I} = \begin{pmatrix} -\mathrm{E}(\dfrac{\partial^2 \log L}{\partial \lambda^2}) & -\mathrm{E}(\dfrac{\partial^2 \log L}{\partial \lambda \partial \beta}) \\ -\mathrm{E}(\dfrac{\partial^2 \log L}{\partial \lambda \partial \beta}) & -\mathrm{E}(\dfrac{\partial^2 \log L}{\partial \beta^2}) \end{pmatrix}.$$

This simplifies to

$$\mathcal{I} = \begin{pmatrix} \dfrac{nm}{\lambda^2} & \displaystyle\sum_{i=1}^{n} \mathrm{E}(T_{i,m}^{\beta} \log T_{i,m}) \\ \displaystyle\sum_{i=1}^{n} \mathrm{E}(T_{i,m}^{\beta} \log T_{i,m}) & \dfrac{nm}{\beta^2} \displaystyle\sum_{i=1}^{n} \mathrm{E}(T_{i,m}^{\beta} (\log T_{i,m})^2) \end{pmatrix}.$$

Using results derived in Crow (1974) and Gaudoin (2004), we have

$$\sum_{i=1}^{n} \mathrm{E}(T_{i,m}^{\beta} \log T_{i,m}) = \frac{nm}{\lambda\beta}[\psi(m+1) - \log\lambda],$$

and

$$\sum_{i=1}^{n} \mathrm{E}(T_{i,m}^{\beta} \log^2 T_{i,m}) = \frac{nm}{\lambda\beta^2}[\psi^{(1)}(m+1) + (\psi(m+1) - \log\lambda)^2],$$

where $\phi(z) = \partial \log\Gamma(z)/\partial z$ is the digamma function and $\phi^{(1)}(z) = \partial\phi(z)/\partial z$ is the polygamma function of order 1. By the equivalency of $\phi(m)$ with $\log m$, and $\phi^{(1)}(m)$ with $1/m$, the information matrix can be inverted to $\mathcal{I}^{-1}$ in (46) from the theorem.

CHAPTER V

AN ADJUSTED EMPIRICAL LIKELIHOOD WITH ESTIMATING EQUATION APPROACH FOR MODELING HEAVILY CENSORED ACCELERATED LIFE TEST DATA

5.1 Introduction

In evaluating the reliability of durable products, accelerated life testing (ALT) is commonly applied by stressing specimens at harsher conditions than in normal-use, thereby hastening failure time in tests with short duration. Regression models of replicated data at several stress levels are built to provide extrapolated estimates of lifetime quantities (e.g., 5th or 10th percentile, mean, variance and lifetime distribution) in the normal-use condition for warranty management, product improvement and risk analysis. For newer products where the physics supporting regression models is not clearly understood for extrapolation, the stress levels are usually set closer to the normal-use condition. Because high durability of products and limited testing time, this practice results in heavily censored data. For example, in Meeker and LuValle (1995), tests of printed circuit boards revealed that 68.5% of the data in the lowest stress level are censored after 4,078 hours (169.9 days) of testing. This creates challenges in deriving statistical inference procedures for lifetime quantities.

Various parametric approaches have been introduced to solve this inference problem. Typical parametric approaches assume that failure time distributions under various stress levels belong to the same parametric family and there is a (transformed) linear regression structure of the location parameters of these distributions. Most ALT procedures assume a constant variance. There are some exceptions, such as Meeter and Meeker (1994), where it is assumed that the logarithm for each of the scale parameters has a linear regression relationship. According to the research in Hutton and Monaghan (2002) and Pascual and

Montepiedra (2005), selecting an inappropriate lifetime distribution could have significant impact (in terms of estimation bias and variance). However, in data exploration, it is often that several lifetime distributions (e.g., lognormal and Weibull) and are not rejected from goodness-of-fit tests.

In some cases, the traditional ALT models cannot accurately represent the failure time data; the commonly used acceleration function for regression might not be suitable. For example, Meeker and LuValle (1995) used chemical-kinetic knowledge to derive an intricate failure time model which does not fit into the ALT model structure. Because the traditional regression-over-the-mean approach is questionable (especially in this case that the means might not exist), Meeker and LuValle constructed log-linear regression models based on two key chemical-reaction parameters found in differential equations that characterize the failure evolution processes. Although this physics-based approach provides a well-justified ALT model, explicit physical relations are rarely available to aid the data modeling so directly. Thus, there is a need of developing a *data exploration* approach to entertain potential regression models and to examine the goodness-of-fit of the assumed lifetime distribution. For example, the regression relationships between percentiles are used in Section 2 for exploring models.

This article focuses on semiparametric approaches, which drop the assumption of distribution forms, but retain the functional relationship of lifetime distributions at different stress levels. The widely used Cox Proportional Hazards (PH) model (Kalbfleisch and Prentice, 1980) assumes that hazard rates under different stresses are proportional to a baseline hazard rate. The Accelerated Life Test model assumes that the logarithm of the survival functions under different stresses differ only in location parameters. For instance, Yang (1999) assumed that distributions pertaining to various stress levels differ only by a median shift. Recently, the proportional odds model has been proposed as a generalization of PH model and ALT model. Cheng *et al.* (1997) proposed an estimating equation method to estimate the regression parameter $\boldsymbol{\theta}$, and studies its large sample properties. Murphy *et al.* (1997) considered the nonparametric maximum likelihood method to estimate the parameters in the proportional odds model. The asymptotic variances for regression coefficients

are given based on large sample approximation.

Empirical likelihood was developed by Owen (1990) as a general nonparametric inference procedure. Qin and Lawless (1994) demonstrated that the empirical likelihood method with additional estimating equations can be useful in incorporating distribution knowledge to improve estimation quality. Empirical likelihood combines the reliability of nonparametric methods with effectiveness of likelihood methods (Owen, 2001). Furthermore, the confidence regions are automatically determined without estimating the variance of test statistics, which can be difficult in the case of the rank-based regression estimators in censored ALT models. In particular, Lu, Chen and Gan (2002) showed that the EL-EE approach is a natural extension of both Generalized Estimating Equations (GEE; Liang and Zeger, 1986) and Quasi-Likelihood Estimation (QLE) approaches (Wedderburn, 1974) by allowing censored data. Recently, the empirical likelihood method has been shown to work well in difficult inference problems involving censored or truncated data (Pan and Zhou, 2000, Murphy and Van der Vaart, 1997). However, the general form of constraints $\mathrm{E}G(t, \boldsymbol{\theta}) = 0$ has not been considered in the case of right censoring. Chen, Lu and Lin (2005) considered interval-censored data and derivations based on empirical-process theory was not involved. In the case of random censoring, the derivation of asymptotic properties is much more difficult than in the complete sample situation (Qin and Lawless, 1994, Owen 1990) since the intermediate optimal solutions do not have an explicit form.

In this article, the regression relationships are treated as estimating equations (EE) serving as constraints in maximizing the Empirical Likelihood (EL). In the accelerated life test, we are often interested in lower percentiles of lifetime at the normal use condition. So, we construct the estimating equations on lower percentiles. Chen and Hall (1993) have shown that the inference of percentiles can be improved by smoothing the estimating equations. Whang (2003) extend the results to censored percentile regression models. This chapter further studies the empirical likelihood inference for general estimating equations with random censoring, and also considers the asymptotic properties for survival functions.

Section 2 shows some data exploration studies based on two reliability data sets from industrial tests. Section 3 defines the empirical likelihood and formulates the ALT regression

model in the estimating equations. Then, the Adjusted Maximum Empirical Likelihood Estimation (AMELE) method is proposed. Section 4 shows the asymptotic properties for the proposed estimators. More examples of ALT data and asymptotic-efficiency studies are presented in Section 5 to illustrate and compare the proposed methods with parametric MLE. Section 6 provides the conclusion and future work.

5.2 Industrial Testing Examples

We employ two different examples in this section based on circuit board testing and reliability for switches. Our main focus is on a conventional ALT experiment (Example 1). The second example illustrates how life testing can be important in reliability improvement studies. Besides the limited failure time data, other source of information such as degradation data and supplemental knowledge about the physical system could be helpful in setting up estimating equations in the proposed EL-EE method.

EXAMPLE 1. Meeker and LuValle (1995) reported on an accelerated life test experiment for 72 or so printed-circuit-boards (PCB) at four high relative humidity (RH) conditions: 49.5% RH, 62.8% RH, 75.4% RH and 82.4% RH. They integrated chemical kinetics into a probability model to derive the failure time distribution of PCBs at the normal-use conditions (10% and 20% RH). Their data exploration showed that the failure time distribution for data collected at the highest stress level (82.4% RH) seemed to be different from those at other lower stress levels. As stated in their paper, the physics of PCB failure at the 82.4% RH level is not well understood. Since it is the furthest from the normal-use condition, the 72 failure data collected at 82.4% RH level were discarded. Figure 24 shows the Weibull probability plot of the data from three other stress levels. The curvature in the plot indicates that the Weibull lifetime distribution does not adequately fit these data. Note that there are only 22 (out of 70) failures in the lowest stress level with 68.6% of data censored after 169.9 days of testing.

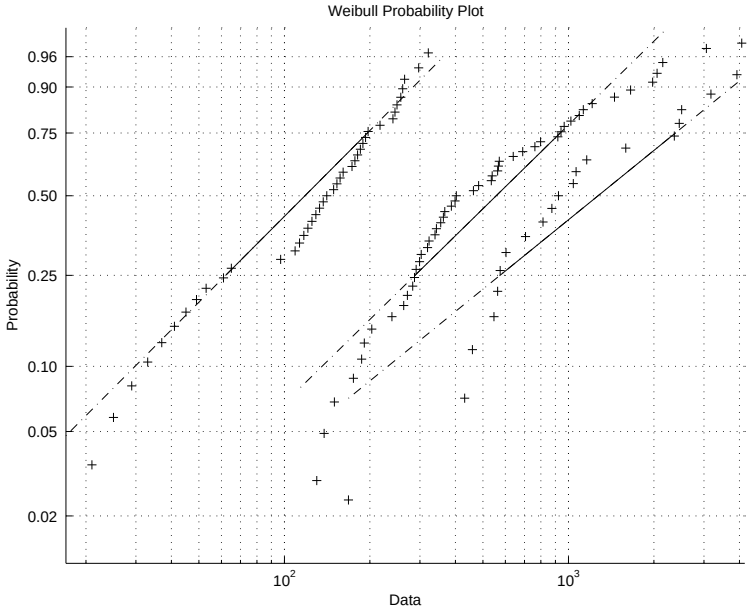

Figure 24: Weibull Probability Plot for the Failure Data Under RH = 49.5, 62.8, 75.4% (from right to left)

The chemical-kinetics-probability distribution derived in their paper is

$$\begin{aligned} Pr(T < t) &= F_T(t; \beta_0^{[k_1]}, \beta_1^{[k_1]}, \beta_0^{[k_2]}, \beta_1^{[k_2]}, \sigma) \\ &= \Phi\left\{\left[-\log[[(k_1+k_2)/k_1]\{1-\exp[-(k_1+k_2)t]\}^{-1}-1]+6\right]/\sigma\right\} \end{aligned}$$

and includes only a tiny proportion (less than 1%) of the population failing at the normal-use condition based on estimated model parameters. In fact, the mean (and variance) of the distribution are not finite, which makes the mean-variance based warranty studies impossible. The following humidity relationships are specified as:

$$k_1 = \exp[\beta_0^{[k_1]} + \beta_1^{[k_1]} g(H)], \quad k_2 = \exp[\beta_0^{[k_2]} + \beta_1^{[k_2]} g(H)],$$

where $\Phi(.)$ is the standard normal cumulative distribution function (cdf) and $g(H) = logit(RH/100)$.

Since lower percentile lifetime such as 5% is observable for all stress levels and important in reliability applications, we explore the regression structures based on them. Figure 25 shows that a simple log-linear regression model is suitable for the EL-EE approach. The proposed AMELE estimation elaborated in Section 5 shows that the estimated 5th

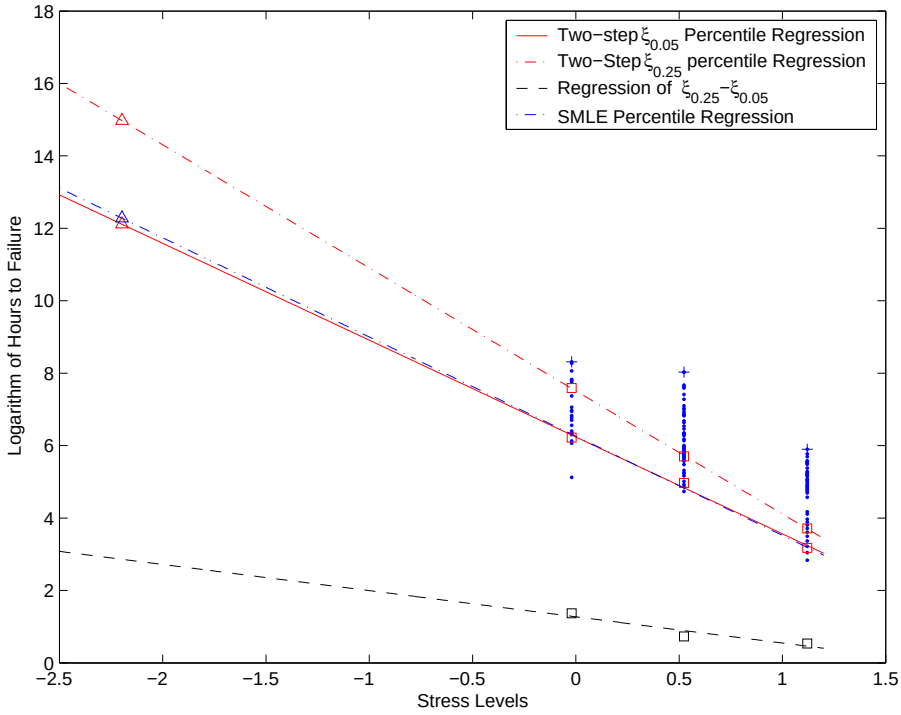

Figure 25: Percentile Regression and Prediction

percentile lifetime of PCBs at 10% RH is 25 years, where the regression prediction based on sample 5th percentiles is 21 years (see the dashed line with shot dots in Figure 2 for the regression). Because Meeker and LuValle's model will predict infinity for the 5th percentile lifetime, it cannot be used as a comparison measure for reliability improvement studies. The percentiles obtained from this data exploration give a more reasonable prediction for comparison measures. The examples presented below reinforce this finding. The width of the confidence interval constructed for the AMELE is 44% shorter than the one deduced from Weibull regression. Section 5 contains a more detailed comparison of the confidence intervals.

Because 25th percentiles are also available for all stress levels, we explore the trend of the "quantile-ranges" (QR) of the logarithm of the 25th and 5th percentiles over three stress levels. Figure 25 shows that the QR is not constant, but rather a linear function with much larger QR in the normal-use condition. For estimating the lifetime distribution, one approach is to assume that after a proper "re-scaling" of the data using the percentile and

QR, the lifetime distributions at all stress levels would be the same. Then, the AMELE gives the estimate and its point-wise confidence intervals. See Section 5 for an example.

EXAMPLE 2. Joseph and Yu (2005) developed robust parameter design methodology for improving product reliability based on degradation data. Their example of a window wiper switch experiment (Wu and Hamada, 2000, page 560) consists of one four-level factor and four two-level factors in an eight-run $OA(4^1 2^4)$ design with four products tested in each run. For each switch, the initial voltage drop (degradation measure) across multiple contacts is recorded every 20,000 cycles up to 180,000. Consider the "soft-failure" definition (Su, Lu, Chen and Hughes-Oliver, 1999) as a drop of voltage over 120. Table 7 presents the failure times. Note that Run #6 has only one failure. Table 2 presents degradation records for the censored cases, which are useful in predicting failure time. Focus on factor effects defined as the difference between the 12.5th percentile life at the high (+1) and low (-1) levels of each factor. Following procedures similar to those used in Joseph and Yu (2005), we identify D and E as significant effects with large differences and we obtained the following percentile regression model based on these differences:

$$y_{12.5\%-tile} = 5.33 - 0.46D - 1.36E.$$

Table 9 reports the failure time data organized in the D and E factor levels useful in the proposed semiparametric method for updating the percentile regression model and estimating the failure time distribution. This table shows that the combination of (D, E) at $(-1, -1)$ gives the largest 12.5th percentile lifetime (the first observation in the table).

EXAMPLE 3. From the empirical likelihood method explained in Section 3, we will find that when the data are heavily censored, there are insufficient failure time points for estimating probability mass. This makes the estimation of the failure time distribution especially troublesome. This example extends Example 2 by exploring ways to handle problems caused by heavily censoring.

Similar to the two effects identified in Joseph and Yu's (2005) mean regression model, B and E are the significant effects for median lifetimes. Run #4 and #6 in Table 7 show

Table 7: Lifetime of Wiper Switches

Run	Factor					Failure Time* of Replicates			
	A	B	C	D	E	1	2	3	4
1	0	−	−	−	−	7.54	8.44	8.73	+(10.52)
2	0	+	+	+	+	4.10	4.69	5.31	8.37
3	1	−	−	+	+	3.89	4.45	4.45	6.99
4	1	+	+	−	−	8.70	+(10.56)	+(12.50)	+(36.75)
5	2	−	+	−	+	4.05	6.46	8.59	9.01
6	2	+	−	+	−	8.75	+(21.56)	+(22.81)	+(26.74)
7	3	−	+	+	−	5.85	6.46	7.07	7.35
8	3	+	−	−	+	7.43	8.65	9.82	+(12.64)

∗: the failure time is given in the unit of (time*−1)*20,000 cycles.
+: censored data (and projected failure time).

Table 8: Degradation Data for Censored Cases

Run	Replicate No.	Inspection Time									
		1	2	3	4	5	6	7	8	9	10
1	4	24	30	38	46	57	71	73	91	98	104
4	2	54	51	64	66	78	84	90	93	106	109
4	3	47	54	63	68	70	77	88	86	91	102
4	4	47	45	50	53	58	57	61	55	61	66
6	2	44	50	48	46	55	63	65	71	68	76
6	3	43	44	55	56	58	62	66	66	72	72
6	4	40	46	45	49	55	62	61	61	64	66
8	4	65	68	69	75	79	84	95	96	101	100

Table 9: Reorganized Wiper Switch Testing Data

Factor		Failure Time*							
D	E	1	2	3	4	5	6	7	8
−	−	7.54	8.44	8.70	8.73	+(10.52)	+(10.56)	+(12.50)	+(36.75)
−	+	4.05	6.46	7.43	8.59	8.65	9.01	9.82	+(12.64)
+	−	5.85	6.46	7.07	7.35	8.75	+(21.56)	+(22.82)	+(26.74)
+	+	3.89	4.10	4.45	4.45	4.69	5.31	6.99	8.37

there is only one failure time and all of the other seven cases are censored. Since these two runs are from the best combination of B and E (at $+1$ and -1 levels, respectively) leading to the largest median lifetime, it is important to develop a method to estimate the lifetime distribution there; we present two approaches below.

First, one could assume that the lifetime distributions at different factor levels are the same (as done in most of ALT experiments) and use either mean or percentile to adjust the data into the same "distribution-scale" such that all the observed failure time points can be used to construct the empirical likelihood. This approach utilizes only the failure time data and not the degradation data. See Section 5.1 for an example.

Alternatively, when degradation data are available, one can either use the degradation path to "project" the failure time or impute the censored data based on the failure time distribution derived from the degradation model. An example for projecting the failure time is to extend the linear regression line of the degradation path for reaching the threshold defining the failure. Sometimes, the long-range extrapolation in ALT (e.g., see Figure 2) causes concern in projecting the failure time. However, if there exists physical knowledge supporting the derivation of failure time distribution, it undoubtedly improves the semiparametric estimation. This example will illustrate how to derive the failure time distribution in the case of heavily censoring. Then, we can combine this knowledge with the observed failure time data at other factor levels for semiparametric estimation.

Consider the following simplification from Joseph and Yu's (2005) model:

$$dY_t/dt = \beta + \sigma W,$$

where Y_t is the voltage degradation drop, t represents the testing-cycle time, and β is the (positive) intercept. The parameter σ is the standard deviation of the white noise W. By integrating with respect to t, we obtain

$$Y_t = Y_0 + \beta t + \sigma W t,$$

where Y_0 is the initial amount of degradation at $t = 0$.

For deriving the failure time distribution, following the traditional random-coefficient

model (e.g., Lu and Meeker, 1993; Lu, Park and Yang, 1997), consider β as a $N(\mu_\beta, \sigma_\beta)$ random variable associated with material (or manufacturing) variations. If W has the standard normal distribution, Y_0 is distributed $N(\mu_{Y_0}, \sigma_{Y_0})$. Denote T the time when degradation reaches the threshold y_f and (soft) failure occurs. Thus,

$$P(T \leq t) = P[(y_f - Y_0)/(\beta + \sigma W)] \approx P[y_f \leq Y_0 + (\beta + \sigma W)t],$$

where the approximation came from the assumption that $P[(\beta + \sigma W) \leq 0] \approx 0$. From aggregating the correlated normal random variables, the cdf of T can be expressed as

$$F_T(t) = 1 - \Phi\left\{\frac{\mu_{Y_0} + \mu_\beta t - y_f}{\sqrt{{\sigma_{Y_0}}^2 + \sigma_\beta{}^2 t^2 + 2t\rho_{Y_0,\beta}\sigma_{Y_0}\sigma_\beta + \sigma^2}}\right\},$$

where we assume the independence between the noise W, β and Y_0, but with a possible correlation between β and Y_0 due to the material property of products.

This model is similar to the failure time cdf derived in Lu, Park and Yang (1997) as a generalization of Berstein's model (Gertsbakh and Kordonskiy, 1969 page 88), where the moments do not exist. However, the percentile lifetime is available for data exploration analysis. For example, the estimates of the parameters in $F_T(t)$ for products tested in Run #6 are given as $\mu_{Y_0} = 46.52$, $\sigma_{Y_0} = 7,37$, $\mu_\beta = 4.25$, $\sigma_\beta = 1.95$ and $\rho_{Y_0,\beta} = 0.98$. Thus, the median lifetime for Run #6 is "derived" as 17.29 in the failure time unit reported in Table 1 and 3. This failure time is equivalent to 325,788 cycles in the experimental setup.

Following the hybrid of parametric and nonparametric maximum likelihood estimation studied in Qin (2000) and Wang, Lu and Kvam (2005), the physics-based failure time distribution for the low-stress-level degradation data can be combined with the empirical likelihood for the high-stress-level failure time data to estimate lifetime in normal-use conditions.

5.3 The Adjusted Empirical Likelihood Estimation Methods

5.3.1 The Empirical Likelihood and ALT Regression Model

Suppose that the accelerated life test is conducted under m different stress levels $\{x_1, ..., x_m\}$, and there are n_j replicates at stress level x_j. Denote the normal-use stress level by x_0. For the jth sample (with stress level x_j), let T_j and C_j be the general notation for the failure

time and censoring random variables, respectively. Assume that C_j and T_j are independent. Denote the probability density function (pdf), the survival function (SF) and the cdf for the failure time T_j as $f_{T_j}(t)$, $S_{T_j}(t) \equiv \Pr(T_j > t) = \int_t^\infty f_{T_j}(x)dx$, $F_{T_j}(t) = 1 - S_{T_j}(t)$, respectively. Similarly, let $f_{C_j}(t)$, $S_{C_j}(t)$ and $F_{C_j}(t)$ be the pdf, SF and cdf of the censoring time C_j.

The general discussion of empirical likelihood for incomplete samples is available from Owen (2003, Chapter 6). The likelihood function is:

$$L = \prod_{j=1}^{m} \left\{ \prod_{i=1}^{k_j} P_{ij} \prod_{i=1}^{k_j+1} \left(\sum_{l=i}^{k_j+1} P_{lj} \right)^{c_{ij}} \right\}, \tag{49}$$

where, for $i = 1, 2, \ldots, k_j, \quad j = 1, 2, \ldots, m$,

$$\begin{aligned} P_{ij} &= \Pr(t_{i-1,j} < T_j \le t_{ij}) = S_{T_j}(t_{i-1,j}) - S_{T_j}(t_{ij}) \\ P_{k_j+1,j} &= \Pr(T > t_{k_j,j}) = S_T(t_{k_j,j}) = 1 - \sum_{i=1}^{k_j} P_{ij}. \end{aligned}$$

Note that, for stress level x_j, the P_{ij} $(i = 1, 2, \ldots, k_j + 1)$ define an empirical distribution with a point mass at each failure time t_{ij}, $i = 1, 2, \ldots, k_j$, $j = 1, 2, \ldots, m$.

The standard nonparametric MLE (NPMLE) of P_{ij}, $i = 1, 2, \ldots, k_j + 1$ can be obtained separately for each sample by finding P_{ij}'s to maximize (49) subject to the constraints $\sum_{i=1}^{k_j+1} P_{ij} = 1$ and $P_{ij} \ge 0$, $i = 1, 2, \ldots, k_j + 1$. With the estimates of P_{ij}'s, we can obtain the estimate of $F_{T_j}(t)$ and $S_{T_j}(t)$. The empirical likelihood approach will maximize the empirical likelihood under the constraints of the m sets of estimating equations,

$$\mathrm{E}[G_j(T_j, \boldsymbol{\theta})] = \sum_{i=1}^{k_j+1} P_{ij} G_j(t_{ij}, \boldsymbol{\theta}) = 0, \tag{50}$$

where $j = 1, 2, \ldots, m$, and $\boldsymbol{\theta}$ is a p-dimensional vector of the regression parameters. Each set of estimating equations could include r functions, $G_j(T, \boldsymbol{\theta}) = (g_{1j}(T_j, \boldsymbol{\theta}), \ldots, g_{rj}(T_j, \boldsymbol{\theta}))^\top$, where $j = 1, 2, \ldots, m$. Thus, there are $s = r \times m (\ge p)$ independent functions to estimate p parameters. For accelerated life tests, a large proportion of observations are being censored, especially in higher stress levels. After a proper transformation, assume a linear regression relationship is plausible for lower percentiles of transformed failure time T_j. We model the

$100q$th percentile of T_j as $x_j^\top \boldsymbol{\theta}$, and then specify the structural relationship in estimating functions:

$$G_j(T_j, \boldsymbol{\theta}) = I(T_j < x_j^\top \boldsymbol{\theta}) - q, \tag{51}$$

where $j = 1, \ldots, m$, $100q$ is the percentile of the lifetime, and $I(\cdot)$ denotes the indicator function. Since this chapter focuses on the lower percentiles in the accelerated life test experiments, we regard (51) as our main constraint functions.

Remark: In situations where censoring is less pronounced, estimating equations can be constructed on the distribution moments. As a simple example, a function of means $\mathrm{E}(T_j) = \psi(x_j, \boldsymbol{\beta})$ is commonly used in ALT regression, where ψ is known (usually, a linear or log-linear function). In this case, $r = 1$ and $m > p$, and

$$\boldsymbol{G}(T, \boldsymbol{\beta}) = (T_1 - \psi(x_1, \boldsymbol{\beta}), T_2 - \psi(x_2, \boldsymbol{\beta}), \ldots, T_m - \psi(x_m, \boldsymbol{\beta}))^\top .$$

Stronger assumptions on the estimating equations can be used to improve the estimation quality. For example, higher moment assumptions can be incorporated via $\boldsymbol{G}$ as

$$\boldsymbol{G}_v(T_j, T_{j'}, x_j, x_{j'}, \boldsymbol{\beta}) = \left(g(T_j, x_j, \boldsymbol{\beta})^v - g(T_{j'}, x_{j'}, \boldsymbol{\beta})^v\right)^\top, \ (j, j') \in \{1, 2, \ldots, m\},$$

where v could be any real number and $g(T_j, x_j, \boldsymbol{\beta})$ can be $T_j - \psi(x_j, \boldsymbol{\beta})$ in the traditional ALT or

$$g(t, x, \boldsymbol{\beta}) = -\log\left(\beta_1 x^{\beta_2} + \frac{\beta_3 x^{\beta_4}}{1 - \exp(-\beta_1 x^{\beta_2}(1 + \beta_3 x^{\beta_4})t)}\right)$$

as derived in Meeker and LuValle (1995).

5.3.2 Adjusted Maximum Empirical Likelihood Estimator

Notice that the constraint functions G in (51) are not differentiable at $\boldsymbol{\theta}$. As shown in Qin and Lawless (1994) and our asymptotics derivation in the following sections, the non-differentiable function G causes problems to theoretical derivation of asymptotic properties. Similar to the techniques proposed in Chen and Hall (1993), a smooth function is introduced to approximate the indicator function. Let K denote an rth-order kernel that is bounded

and compactly supported on [-1,1], satisfying

$$\int u^j K(u)du = \begin{cases} 1 & , \quad \text{if } j = 0, \\ 0 & , \quad \text{if } 1 \leq j \leq r-1, \\ \kappa & , \quad \text{if } j = r, \end{cases} \tag{52}$$

where $r \geq 2$ and constant $\kappa \neq 0$. Define $\mathcal{K}(x) = \int_{y<x} K(y)dy$, and $\mathcal{K}_h(x) = \mathcal{K}(x/h)$. Then, we can have a smoothed version of constraint function G_j, given by

$$G_{j,h} = \mathcal{K}_h(x_j^\top \boldsymbol{\theta}) - T_j) - q, \tag{53}$$

where $j = 1, \ldots, m$.

Let $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset I\!R^p$ be the p-dimensional parameter vector in the estimating equations, where $\boldsymbol{\Theta}$ is the parameter space containing a neighborhood of the true parameter $\boldsymbol{\theta}_0$. Given $\boldsymbol{\theta}$, the optimal P_{ij} can be solved by maximizing empirical likelihood L in (49) subject to the constraint in (53). Using the standard Lagrange multiplier arguments as in Owen (1990), Qin and Lawless (1994) and Owen (2001), we can have the following intermediate results:

$$\begin{aligned} \hat{P}_{ij}(\boldsymbol{\lambda}_j) &= \frac{1}{n_j(1 - a_{ij}(\boldsymbol{\lambda}_j) + \boldsymbol{\lambda}_j^\top G_{j,h}(t_{ij}, \boldsymbol{\theta}))}, \\ \hat{P}_{k_j+1,j}(\boldsymbol{\lambda}_j) &= 1 - \sum_{i=1}^{k_j} \hat{P}_{ij}(\boldsymbol{\lambda}_j), \end{aligned} \tag{54}$$

where

$$a_{ij}(\boldsymbol{\lambda}_j) = \frac{1}{n_j} \sum_{v=1}^{i} \frac{c_{vj}}{\sum_{l=v}^{k_j+1} \hat{P}_{lj}(\boldsymbol{\theta})}, \tag{55}$$

is an implicit function of $\boldsymbol{\lambda}_j$ and $\boldsymbol{\theta}$, and $\boldsymbol{\lambda}_j$ is the solution to

$$\sum_{i=1}^{k_j+1} \frac{G_{j,h}(t_{ij}, \boldsymbol{\theta})}{1 - a_{ij}(\boldsymbol{\lambda}_j) + \boldsymbol{\lambda}_j^\top G_{j,h}(t_{ij}, \boldsymbol{\theta})} = 0, \quad j = 1, 2, \ldots, m. \tag{56}$$

Remark: When there are no constraints G_{jh}, the optimal $\hat{P}_{ij}$ maximizing the nonparametric likelihood is the Kaplan-Meier estimator (Kaplan and Meier, 1958). By setting $\boldsymbol{\lambda}_j$=0, $P_{ij}(0) = [n_j(1 - a_{ij}(0))]^{-1}$. The equivalence of $P_{ij}(0)$ and the Kaplan-Meier estimator

$d\hat{F}_{T_j,KM}(T_i)$ is shown in Appendix A. In the case that there is only one sample and no censoring is observed, i.e., $a_{ij} = 0$, and $m = 1$, equations (54) and (56) reduce to

$$P_i = \frac{1}{n(1 + \boldsymbol{\lambda}^\top G(t_i, \boldsymbol{\theta}))}, 1 \le i \le n, \quad \text{and} \quad \sum_{i=1}^{n} \frac{G(t_i, \boldsymbol{\theta})}{1 + \boldsymbol{\lambda}^\top G(t_i, \boldsymbol{\theta})} = 0,$$

as shown in Qin and Lawless (1994).

For each $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset I\!R^p$, we can find corresponding optimal $\boldsymbol{\lambda}_j$ and $\hat{P}_{ij}(\boldsymbol{\lambda}_j)$ to maximize the nonparametric likelihood defined in (49). Plugging $\hat{P}_{ij}(\boldsymbol{\lambda}_j)$ back into likelihood (49), we have a profile likelihood of $\boldsymbol{\theta}$ as $L(\boldsymbol{\theta})$. The maximum empirical likelihood estimator of $\boldsymbol{\theta}$ can then be solved by maximizing $\boldsymbol{\theta}$ over parameter space $\boldsymbol{\Theta}$. However, according to intermediate results (54),(55),(56), there is no explicit solution available for $\hat{P}_{ij}(\boldsymbol{\lambda}_j)$. The computations for solving optimal values become quite complicated. Moreover, due to the interdependence of $P_{ij}(\boldsymbol{\lambda}_j)$, $\boldsymbol{\lambda}_j$ and $a_{ij}(\boldsymbol{\lambda}_j)$, the asymptotic property of the optimal estimator is hard to obtain. Here we circumvent this difficulty by replacing $1 - a_{ij}(\boldsymbol{\lambda}_j)$ by $1 - a_{ij}(0)$. Therefore, for given $\boldsymbol{\theta}$, the intermediate results in (54),(55),(56) become:

$$\begin{aligned} \tilde{P}_{ij}(\boldsymbol{\lambda}_j) &= \frac{1}{n_j(1 - a_{ij}(0) + \boldsymbol{\lambda}_j^\top G_{j,h}(t_{ij}, \boldsymbol{\theta}))}, \\ \tilde{P}_{k_j+1,j}(\boldsymbol{\lambda}_j) &= 1 - \sum_{i=1}^{k_j} \tilde{P}_{ij}(\boldsymbol{\lambda}_j), \end{aligned} \tag{57}$$

where

$$a_{ij}(0) = \frac{1}{n_j} \sum_{v=1}^{i} \frac{c_{vj}}{\sum_{l=v}^{k_j+1} \tilde{P}_{lj}(0)}, \tag{58}$$

and $\boldsymbol{\lambda}_j$ is the solution to

$$\sum_{i=1}^{k_j+1} \frac{G_{j,h}(t_{ij}, \boldsymbol{\theta})}{1 - a_{ij}(0) + \boldsymbol{\lambda}_j^\top G_{j,h}(t_{ij}, \boldsymbol{\theta})} = 0, \quad j = 1, 2, \ldots, m. \tag{59}$$

Plugging the new set of intermediate results into likelihood (49), we can obtain a new estimator $\tilde{\boldsymbol{\theta}}$ by maximizing the profile likelihood. We name this new estimator as Adjusted Maximum Empirical Likelihood Estimator (AMELE), which has some nice properties as shown in the following sections.

5.3.3 Regularity Conditions

Let $n_{max} = max_j\{n_j\}$, $n_{min} = min_j\{n_j\}$. The following regularity conditions are necessary in discussing the properties of the AMELE $\tilde{\boldsymbol{\theta}}$ and $\tilde{S}_{T_j}(t)$.

(**R.1**) Parameter space $\boldsymbol{\Theta} \subset R^p$ is compact and contains a neighborhood of true parameter $\boldsymbol{\theta}_0$, and, for $H(\boldsymbol{\theta})$ given in (63), $|R| < \infty$, where

$$|R| = \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \{0 < |H(\boldsymbol{\theta})|\}.$$

(**R.2**) Given $\boldsymbol{t}_j = (t_{1j}, t_{2j}, \ldots, t_{k_j,j}, t_{k_j+1,j})$, let $\boldsymbol{G}_{j,h}(\boldsymbol{t}_j, \boldsymbol{\theta}) = (G_{j,h}(t_{ij}, \boldsymbol{\theta}))_{(k_j+1)\times r}$. For every $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, assume that $r \times r$ matrix $G_{j,h}^{\top} G_{j,h}$, $j = 1, 2, \ldots, m$, is nonsingular.

(**R.3**) For $j = 1, 2, \ldots, m$, $\mathrm{E}(||G_{j,h}(T, \boldsymbol{\theta})||^3) < \infty$ and $G_j(T, \boldsymbol{\theta})$ is second-order differentiable with respect to $\boldsymbol{\theta}$, i.e., $\partial^2 G_{j,h}(T, \boldsymbol{\theta}) / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\top}$ exists for each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

(**R.4**) Let $F_0(t)$ denote the distribution function of $T - x'\boldsymbol{\theta}$ such that $F_0(0) = q$, then $f_0(u)$ is bounded away from zero, and is r times continuously differentiable with respect to u for all u in a neighborhood of 0.

(**R.5**) $\lim_{n\to\infty} nh^{2r} = 0$.

The regularity condition **R1** is to ensure that the maximum of $|H(\boldsymbol{\theta})|$ exists in the interior of $\boldsymbol{\Theta}$. **R2** and **R3, R4** require the non-singularity, continuity and differentiability of estimating function $G_{jh}(t, \boldsymbol{\theta})$ to ensure that equation (64) is well defined and the AMELE $\tilde{\boldsymbol{\theta}}$ is in $||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| < n_{min}^{-1/3}$ with probability one, given that n_{min} is sufficiently large. **R4** ensures the smoothing parameter h goes to 0 at a proper rate. Note that we don't consider the condition $nh/\log n \to \infty$ as in Chen and Hall (1993), since Edgeworth analysis is not the main focus of this chapter.

For notational simplicity, define

$$\begin{aligned} Z_{ij,h}(\boldsymbol{\theta}) &= \frac{G_{j,h}(t_{ij}, \boldsymbol{\theta})}{1 - a_{ij}(0)} \\ &= G_{j,h}(t_{ij}, \boldsymbol{\theta}) d\hat{F}_{T_j,KM}(t_{ij}), \end{aligned} \tag{60}$$

where the equivalence of $1 - a_{ij}(0)$ and $d\hat{F}_{T_j,KM}(t_{ij})$ is shown in the Appendix A. Then, (57), (59) can be simplified as

$$\begin{aligned} \tilde{P}_{ij}(\boldsymbol{\lambda}_j) &= \frac{1}{1+\boldsymbol{\lambda}_j^\top Z_{ij,h}(\boldsymbol{\theta})} d\hat{F}_{T_j,KM}(t_{ij}), \qquad (61) \\ \tilde{P}_{k_j+1,j}(\boldsymbol{\lambda}_j) &= 1 - \sum_{i=1}^{k_j} \tilde{P}_{ij}(\boldsymbol{\lambda}_j), \end{aligned}$$

$$\sum_{i=1}^{k_j+1} \frac{Z_{ij,h}(\boldsymbol{\theta})}{1+\boldsymbol{\lambda}_j^\top Z_{ij,h}(\boldsymbol{\theta})} = 0, \quad j = 1, 2, \ldots, m. \qquad (62)$$

Similar to Lemma 2 in Whang (2003) regarding the complete sample case, we have the following results for the $Z_{ij,h}$ defined in (61):

Lemma 1 *Under the regularity conditions, as* $n_{min} \to \infty$,

$$\begin{aligned} &(a)\ \frac{1}{n_j} \sum_{i=1}^{k_j+1} Z_{ij,h}(\boldsymbol{\theta}) = \int G_{jh}(t,\boldsymbol{\theta}) d\hat{F}_{T_j,KM}(t) = O_p(n^{-1/2}), \\ &(b)\ \frac{1}{n_j} \sum_{i=1}^{k_j+1} Z_{ij,h}(\boldsymbol{\theta}) Z_{ij,h}(\boldsymbol{\theta})^\top = \boldsymbol{A}_j(\boldsymbol{\theta}_0) + o_p(1), \\ &(c)\ E\left[\frac{\partial G_{jh}(T,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right] = \boldsymbol{B}_j(\boldsymbol{\theta}_0) + o_p(1) \end{aligned}$$

uniformly in the ball $||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| \le n^{-1/2}$

For a given $\boldsymbol{\theta}$, a unique $\boldsymbol{\lambda}_j$ exists, provided that 0 is inside the convex hull of the points $Z_{ij,h}(\boldsymbol{\theta})$. The following lemma quantify the magnitude of the $\boldsymbol{\lambda}_j$ in a small neighborhood of $\boldsymbol{\theta}$.

Lemma 2 *Let* $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ *be the true value of the parameter. Under the regularity conditions* **R1-R4**, *we have the following results: For* $\boldsymbol{\theta} \in \{\boldsymbol{\theta} : ||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| \le n_{max}^{-1/3}\}$ *and* $\boldsymbol{\lambda}_j(\boldsymbol{\theta})$ *satisfying (62), we have* $\boldsymbol{\lambda}_j(\boldsymbol{\theta}) \xrightarrow{w.p.1} 0$, $j = 1, 2, \ldots m$, *and* $\boldsymbol{\lambda}_j(\boldsymbol{\theta}) = O_p(n^{-1/2})$ *uniformly, as* $n_{min} \to \infty$.

Given an unique $\boldsymbol{\lambda}_j(\boldsymbol{\theta})$, the $\tilde{P}_{ij}(\boldsymbol{\lambda}_j)$ is well defined through (61). The adjusted log-profile-likelihood is

$$H(\boldsymbol{\theta}) = \sum_{j=1}^{m} \left\{ \sum_{i=1}^{k_j} \log \tilde{P}_{ij}(\boldsymbol{\theta}) + \sum_{i=1}^{k_j+1} c_{ij} \log \left(\sum_{l=i}^{k_j+1} \tilde{P}_{lj}(\boldsymbol{\theta}) \right) \right\}. \qquad (63)$$

Following the same argument in Di, Lu, and Lin (2003), the adjusted likelihood equation $\partial H(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = 0$ can be simplified to

$$\begin{aligned}\frac{\partial H(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} &= -\sum_{j=1}^{m}\left[n_j\boldsymbol{\lambda}_j^{\top}(\boldsymbol{\theta})\sum_{i=1}^{k_j}\frac{\partial G_{jh}(t_{ij},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\tilde{P}_{ij}(\boldsymbol{\theta})\right] \\ &= -\sum_{j=1}^{m}\left[n_j\boldsymbol{\lambda}_j^{\top}(\boldsymbol{\theta})\sum_{i=1}^{k_j}\frac{\partial Z_{ijh}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{1}{1+\boldsymbol{\lambda}_j^{\top}(\boldsymbol{\theta})Z_{ij,h}(\boldsymbol{\theta})}\right] \\ &= 0.\end{aligned} \tag{64}$$

Denote $\tilde{\boldsymbol{\theta}}$ the solution of equation (64) as the AMELE for the parameter $\boldsymbol{\theta}$, and the corresponding AMELE for the survival function $S_{T_j}(t)$ can be written as

$$\tilde{S}_{T_j}(t) = \sum_{t_{ij}>t}\frac{1}{n_j\left(1-a_{ij}(0)+\boldsymbol{\lambda}_j^{\top}G_{jh}(t_{ij},\tilde{\boldsymbol{\theta}})\right)}, \tag{65}$$

where $a_{ij}(0)$ and $\boldsymbol{\lambda}_j$ are defined by (54) - (56) with respect to $\tilde{\boldsymbol{\theta}}$.

The following lemma justifies the AMELE $\tilde{\boldsymbol{\theta}}$ as a consistent estimator of $\boldsymbol{\theta}_0$, which can be proved using the same arguments in Qin and Lawless (1994).

Lemma 3 *Under the regularity conditions, as $n_{min} \to \infty$,*

(a) There exist $\tilde{\boldsymbol{\theta}} \in \boldsymbol{\Theta}$ and $\boldsymbol{\lambda}_j = \boldsymbol{\lambda}_j(\tilde{\boldsymbol{\theta}})$ satisfying equations (64) and (62).

(b) When n_{min} is large, $H(\boldsymbol{\theta})$ attains its maximum value at some point $\tilde{\boldsymbol{\theta}}$ in the interior of the ball $||\boldsymbol{\theta}-\boldsymbol{\theta}_0|| \le n_{max}^{-1/3}$ with probability one. Thus, the AMELE $\tilde{\boldsymbol{\theta}}$ is a strongly consistent estimate of $\boldsymbol{\theta}_0$.

□

5.4 Asymptotic Properties of the AMELE

In the investigation of the asymptotic properties of AMELE, we start with understanding the large sample properties of $\boldsymbol{\lambda}_j(\tilde{\boldsymbol{\theta}})$.

5.4.1 Asymptotic Distribution of $\boldsymbol{\lambda}_j$

For $\boldsymbol{\theta} \in \{\boldsymbol{\theta} : ||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| \leq n^{-1/2}\}$, $\boldsymbol{\lambda}_j = O_p(n^{-1/2})$ according to Lemma 2. If we expand (62) at $\boldsymbol{\lambda}_j = 0$, we have

$$\begin{aligned} \boldsymbol{\lambda}_j &= -\left[\frac{1}{n_j}\sum_{i=1}^{k_j+1} Z_{ij,h}(\boldsymbol{\theta}) Z_{ij,h}^{\top}(\boldsymbol{\theta})\right]^{-1}\left[\frac{1}{n_j}\sum_{i=1}^{k_j+1} Z_{ij,h}(\boldsymbol{\theta})\right] + o_p(n^{-1/2}) \qquad (66) \\ &= -\boldsymbol{A}_j(\boldsymbol{\theta}_0)^{-1}\int G_{jh}(t,\boldsymbol{\theta}) d\hat{F}_{T_j,KM}(t) + o_p(n^{-1/2}), \end{aligned}$$

where the second equation follows from Lemma 1.

We use the following Theorem to state the asymptotic normality of $\boldsymbol{\lambda}_j$.

Theorem 1 *For continuous lifetime T_j and censoring time C_j, suppose $S_{C_j}(L) > 0$, and $S_{T_j}(t)$ is continuous at $t = L_j$. Then, as $n_{min} \to \infty$, if $\boldsymbol{\theta} \in \{\boldsymbol{\theta} : ||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| \leq n_{max}^{-1/3}\}$,*

$$\sqrt{n_j}\boldsymbol{\lambda}_j(\boldsymbol{\theta}) \xrightarrow{d} N_r(0, \boldsymbol{\Sigma}_{\lambda_j}(\boldsymbol{\theta})), \quad j = 1, 2, \ldots m,$$

where

$$\boldsymbol{\Sigma}_{\lambda_j}(\boldsymbol{\theta}) = \boldsymbol{A}_j(\boldsymbol{\theta})^{-1}\boldsymbol{\Sigma}_{G_j}(\boldsymbol{\theta})\boldsymbol{A}_j(\boldsymbol{\theta})^{-1}, \qquad (67)$$

$$\boldsymbol{\Sigma}_{G_j}(\boldsymbol{\theta}) = \int_0^{\infty}\left\{G_j(t,\boldsymbol{\theta})[1 - F_{T_j}(t)] - \int_0^t G_j(s,\boldsymbol{\theta}) dF_{T_j}(t)\right\}^2 \frac{dF_{T_j}(t)}{[1 - F_{T_j}(t)]^2[1 - F_{C_j}(t)]}, \qquad (68)$$

and $\boldsymbol{A}_j(\boldsymbol{\theta})$ is given in (79).

Considering (64), denote

$$\boldsymbol{l}(\boldsymbol{\theta}) = \frac{1}{n}\frac{\partial H(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{j=1}^{m}\left\{\frac{n_j}{n}\boldsymbol{\lambda}_j^{\top}(\boldsymbol{\theta})\sum_{i=1}^{k_j+1}\tilde{P}_{ij}(\boldsymbol{\theta})\frac{\partial G_{jh}(t_{ij},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\}$$

the partial derivative of the profile likelihood, where $n = \sum_{j=1}^m n_j$. Considering $\boldsymbol{\lambda}_j \to 0$ with probability 1, and uniform consistency of $\hat{F}_{T,KM}(t)$, we have the following statements

for G_{jh} defined in (53),

$$\begin{aligned}\sum_{i=1}^{k_j+1}\tilde{P}_{ij}(\boldsymbol{\theta})\frac{\partial G_{jh}(t_{ij},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} &= \int\frac{\partial G_{jh}(t,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{1}{1+\boldsymbol{\lambda}_j Z_{ij,h}(\boldsymbol{\theta})}d\hat{F}_{T,KM}(t)\\ &\to \int\frac{\partial G_{jh}(t,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}d\hat{F}_{T,KM}(t)\\ &\to \int\frac{\partial G_{jh}(t,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}dF_T(t)=\mathrm{E}\frac{\partial G_{jh}(t,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\\ &\to \boldsymbol{B}_j(\boldsymbol{\theta}_0),\end{aligned}$$

as $n_{min}\to\infty$, for given $\boldsymbol{\theta}\in\{\boldsymbol{\theta}:||\boldsymbol{\theta}-\boldsymbol{\theta}_0||\le n^{-1/2}\}$. The last equation follows from part(c) in Lemma 1.

From the independence of the m samples, the asymptotic normality of $\boldsymbol{l}(\boldsymbol{\theta})$ follows directly from Theorem 1. We state it as the following corollary.

Corollary 1 *Let $n=\sum_{j=1}^m n_j$, and assume $n_j/n\to\gamma_j$ as $n_{min}\to\infty$, where $0\le\gamma_j\le 1$. Under the conditions of Theorem 1, for given $\boldsymbol{\theta}\in\{\boldsymbol{\theta}:||\boldsymbol{\theta}-\boldsymbol{\theta}_0||\le n_{max}^{1/3}\}$, then $\sqrt{n}\boldsymbol{l}_n(\boldsymbol{\theta})$ is asymptotically normal with mean zero, and covariance matrix*

$$\boldsymbol{\Sigma}_l(\boldsymbol{\theta})=\sum_{j=1}^m\sqrt{\gamma_j}\boldsymbol{B}_j(\boldsymbol{\theta})\boldsymbol{\Sigma}_{\lambda_j}(\boldsymbol{\theta})\boldsymbol{B}_j(\boldsymbol{\theta})^\top, \tag{69}$$

where $\boldsymbol{\Sigma}_{\lambda_j}(\boldsymbol{\theta})$ is given by (67).

5.4.2 Asymptotic Normality of the AMELE of Model Parameters

Applying Taylor's expansion to $\boldsymbol{l}(\boldsymbol{\theta})$ around $\boldsymbol{\theta}_0$, we have

$$0=\boldsymbol{l}(\tilde{\boldsymbol{\theta}})=\boldsymbol{l}(\boldsymbol{\theta}_0)-\frac{\partial\boldsymbol{l}(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)+o_p(\|\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\|). \tag{70}$$

Take partial derivative of $\boldsymbol{l}(\boldsymbol{\theta})$ with regard to $\boldsymbol{\theta}$,

$$\frac{\partial\boldsymbol{l}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}=\sum_{j=1}^m\left\{\frac{n_j}{n}\boldsymbol{\lambda}_j^\top(\boldsymbol{\theta})\frac{\partial}{\partial\boldsymbol{\theta}}\left(\sum_{i=1}^{k_j+1}\tilde{P}_{ij}(\boldsymbol{\theta})\frac{\partial G_{jh}(t_{ij},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)+\frac{n_j}{n}\frac{\partial\boldsymbol{\lambda}_j^\top(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\left(\sum_{i=1}^{k_j+1}\tilde{P}_{ij}(\boldsymbol{\theta})\frac{\partial G_{jh}(t_{ij},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)\right\}.$$

As $n_{min}\to\infty$, the first term of right side in the above equation goes to zero in probability (since $\boldsymbol{\lambda}_j(\boldsymbol{\theta})\xrightarrow{w.p.1}0$), and the convergence of the second part is shown in Lemma 4. Thus,

$$\lim_{n\to\infty}\frac{\partial\boldsymbol{l}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}=\lim_{n\to\infty}\sum_{j=1}^m\frac{n_j}{n}\frac{\partial\boldsymbol{\lambda}_j(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\boldsymbol{B}_j(\boldsymbol{\theta}_0).$$

It follows from (66) that

$$\begin{aligned}\lim_{n\to\infty}\frac{\partial\boldsymbol{\lambda}_j(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} &= \frac{\partial\boldsymbol{A}_j(\boldsymbol{\theta})^{-1}}{\partial\boldsymbol{\theta}}\mathrm{E}(G_{jh}(T_j,\boldsymbol{\theta})) + \boldsymbol{A}_j(\boldsymbol{\theta})^{-1}\boldsymbol{B}_j(\boldsymbol{\theta}) \\ &= \boldsymbol{A}_j(\boldsymbol{\theta})^{-1}\boldsymbol{B}_j(\boldsymbol{\theta}).\end{aligned}$$

Therefore,

$$\lim_{n\to\infty}\frac{\partial\boldsymbol{l}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = \sum_{j=1}^{m}\left\{\gamma_j\boldsymbol{B}_j(\boldsymbol{\theta})^\top\boldsymbol{A}_j(\boldsymbol{\theta})^{-1}\boldsymbol{B}_j(\boldsymbol{\theta})\right\},$$

where $n_j/n \to \gamma_j$. Applying Corollary 4.1 and (70), we have the following theorem.

Theorem 2 *Under the regularity conditions* **R.1 - R.5**, *suppose* $n_j/n \to \gamma_j$ *as* $n \to \infty$, *where* $0 \le \gamma_j \le 1$. *Then,* $\sqrt{n}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N_p(0, \boldsymbol{\Sigma}_\theta)$, *where*

$$\boldsymbol{\Sigma}_\theta = \boldsymbol{C}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{\Sigma}_l(\boldsymbol{\theta}_0)\boldsymbol{C}(\boldsymbol{\theta}_0)^{-1},$$

$$\boldsymbol{C}(\boldsymbol{\theta}_0) = \sum_{j=1}^{m}\left\{\gamma_j\boldsymbol{B}(\boldsymbol{\theta}_0)^\top\boldsymbol{A}_j(\boldsymbol{\theta}_0)^{-1}\boldsymbol{B}(\boldsymbol{\theta}_0)\right\},$$

and $\boldsymbol{A}_j(\boldsymbol{\theta}_0)$ *and* $\boldsymbol{\Sigma}_l(\boldsymbol{\theta}_0)$ *are given by (79) and (69), respectively.*

The following theorem shows that the AMELE has similar likelihood ratio test as the parametric MLE, which be proved following similar procedures used in Qin and Lawless (1994).

Theorem 3 *Under the regularity conditions* **R.1 - R.5**,

(1) When $H_{10} : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ *is true,* $\Lambda_{1n}(\boldsymbol{\theta}_0) = 2(H(\tilde{\boldsymbol{\theta}}) - H(\boldsymbol{\theta}_0)) \xrightarrow{d} \chi^2(p)$, *where* p *is the number of the parameters in* $\boldsymbol{\theta}$, *and* H *is the log-likelihood defined in (63).*

(2) Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, *where* $\boldsymbol{\theta}_1$ *and* $\boldsymbol{\theta}_2$ *are* $q \times 1$ *and* $(p-q) \times 1$ *vectors, respectively. For* $H_{20} : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_{10}$, *the profile empirical likelihood ratio test (PELRT) statistic is*

$$\Lambda_{2n}(\boldsymbol{\theta}_{10}) = 2(H(\tilde{\boldsymbol{\theta}}_1, \tilde{\boldsymbol{\theta}}_2) - H(\boldsymbol{\theta}_{10}, \tilde{\boldsymbol{\theta}}_2)),$$

where $\tilde{\boldsymbol{\theta}}_2$ *maximizes* $H(\boldsymbol{\theta}_{10}, \boldsymbol{\theta}_2)$ *with respect to* $\boldsymbol{\theta}_2$. *Under* H_{20}, $\Lambda_{2n} \xrightarrow{d} \chi^2(q)$, *as* $n \to \infty$.

Part (2) of Theorem 4.5 can be used to test whether the assumptions $\mathrm{E}(G(T,\boldsymbol{\theta}))=0$ are adequate, where $G=\{G_1,G_2,\ldots,G_m\}$. For testing this hypothesis, the PELRT statistic is $\Lambda_{3n}(\boldsymbol{\theta})=2(H^*-H(\tilde{\boldsymbol{\theta}}))$, where H^* is the maximum of the empirical log-likelihood in the NPMLE approach without any constraint (given by the Kaplan-Meier estimate), and $H(\tilde{\boldsymbol{\theta}})$ is the log-likelihood function with $\tilde{\boldsymbol{\theta}}$ as the AMELE estimates.

To derive the asymptotic distribution of $\Lambda_{3n}(\boldsymbol{\theta})$, we first recall that $G_h(t,\boldsymbol{\theta})=(g_{1h}(t,\boldsymbol{\theta}),g_{2h}(t,\boldsymbol{\theta}),\ldots,g_{sh}(t,\boldsymbol{\theta}))$, where $s=r\times m$; and $\boldsymbol{\theta}=(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,\ldots,\boldsymbol{\theta}_p)$. Note that any p of s $(s>p)$ equations $\mathrm{E}(g_{ih}(T,\boldsymbol{\theta}))=0$, $i=1,2,\ldots,s$, can define parameter $\boldsymbol{\theta}=(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,\ldots,\boldsymbol{\theta}_p)$. For convenience, let $\boldsymbol{\theta}=(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,\ldots,\boldsymbol{\theta}_p)$ be defined by the first p equations, i.e., $\mathrm{E}(g_{ih}(T,\boldsymbol{\theta}))=0$, $i=1,2,\ldots,p$. Denote $\beta_j=\mathrm{E}(g_{jh}(T,\boldsymbol{\theta}))$, $j=p+1,p+2,\ldots,s$, and $\tau=(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,\ldots,\boldsymbol{\theta}_p,\beta_1,\beta_2,\ldots,\beta_{s-p})$. Then, the r-dimensional parameter vector τ can be determined by the estimating equations $\mathrm{E}(G_h^*(T,\delta))=0$, where $G_h^*(t,\delta)=(g_{1h}^*(t,\delta),g_{2h}^*(t,\delta),\ldots,g_{sh}^*(t,\delta))$, and $g_{ih}^*(t,\delta)=g_{ih}(t,\boldsymbol{\theta})$, $i=1,2,\ldots,p$, $g_{ih}^*(t,\delta)=g_{ih}(t,\boldsymbol{\theta})-\beta_{i-p}$, $i=p+1,p+2,\ldots,s$.

With this re-parameterization, testing the model $H_{20}:\mathrm{E}(G_h(T,\boldsymbol{\theta}))=0$ is equivalent to testing $H_{20}^*:\boldsymbol{\beta}=0$, where $\boldsymbol{\beta}=(\beta_1,\beta_2,...,\beta_{s-p})$. The PELRT statistic in this case becomes

$$\tilde{\Lambda}_{2n}=2(H(\tilde{\tau})-H(tilde\theta,0)).$$

In the case of $p=s$, the estimating equations $\mathrm{E}(G_h^*(T,\tau))=0$ specify no constraints on τ, and $H(\hat{\tau})$ is equal to H^*, the maximum of the empirical likelihood. With $H_n(\tilde{\theta},0)=H(\tilde{\theta}_n)$, the maximum likelihood in the AMELE approach, we have

$$\tilde{\Lambda}_{3n}=2(H(\tilde{\tau})-H(\tilde{\theta},0))=2(H^*-H(\tilde{\theta}))=\Lambda_{2n}.$$

This leads to the following corollary.

Corollary 2 *If* $\mathrm{E}(G_j(T,\boldsymbol{\theta}))=0$, *then under the regularity conditions* **R.1 - R.3**, *we have* $\Lambda_{3n}(\boldsymbol{\theta})=2(H^*-H(\hat{\boldsymbol{\theta}}_n))\xrightarrow{d}\chi^2(s-p)$, *where* s *is the number of total independent functions specified in* $G_j(T,\boldsymbol{\theta})$, $j=1,2,\ldots,m$, *and* p *is dimension of vector* $\boldsymbol{\theta}$.

The asymptotic properties for the AMELE of the survival function $S_{T_j}(t)$ can be proved in the similar manner with the detailed proof in the Appendix.

Theorem 4 *Under the conditions of Theorem 1, as* $n_{min} \to \infty$, $\sqrt{n_j}(\tilde{S}_{T_j,}(t) - S_{T_j}(t)) \xrightarrow{d}$ $N_m(0, \mathbf{\Sigma}_{S_j(t)})$, *where*

$$\begin{aligned}\sigma^2_{S_j(t)} &= S^2_{T_j}(t)\int_0^t \frac{dF_{T_j}(x)}{S_{T_j}(x)^2 S_{C_j}(x)} + \\ &\int_0^\infty \left[\int_x^\infty \boldsymbol{\beta}_j^\top(t,\boldsymbol{\theta})(G_j(x,\boldsymbol{\theta}) - G_j(t,\boldsymbol{\theta}))dS_{T_j}(t)\right]^2 \frac{dF_{T_j}(x)}{S^2_{T_j}(x)S_{C_j}(x)} + \\ &2S_{T_j}(t)\int_0^t \left(\int_s^\infty \boldsymbol{\beta}_j^\top(t,\boldsymbol{\theta})(G_j(s,\boldsymbol{\theta}) - G_j(x,\boldsymbol{\theta})dS_{T_j}(x)\right)\frac{dF_{T_j}(s)}{S_{C_j}(s)},\end{aligned}$$

$$\boldsymbol{\beta}_j(t,\boldsymbol{\theta}) = A_j(\boldsymbol{\theta})^{-1}\int_t^\infty \frac{G_j(x,\boldsymbol{\theta})dF_{T_j}(x)}{S_{C_j}(x)},$$

and $A_j(\boldsymbol{\theta})$ is given in (79).

5.5 Numerical Evaluation

In the following numerical examples, we choose a kernel function used in Whang (2003) for illustration purpose. The $\mathcal{K}(u)$ is the integral of a fourth-order kernel given by

$$\mathcal{K}(u) = \begin{cases} 0, & \text{if } u < -1 \\ 0.5 + \frac{105}{64}[u - \frac{5}{3}u^3 + \frac{7}{5}u^3 - \frac{3}{7}u^7], & \text{if } |u| \le 1 \\ 1, & \text{otherwise.} \end{cases} \tag{71}$$

The smoothing parameter h is chosen as 0.1 for a smooth visualization of the profile likelihood.

5.5.1 Asymptotic Bias and Variance Studies

Consider interval-censored data. Chen, Lu and Lin (2005) investigated the asymptotic efficiency of the empirical likelihood estimates versus the parametric MLE for mis-specified lifetime distributions. They concluded that if the distribution assumption is correct for the parametric MLE, the performance of the empirical likelihood estimates is still close to the MLE with more than 80% asymptotic efficiency. When the distribution assumption is incorrect, the asymptotic Mean Squared Error (MSE) is less than 10% of asymptotic MSE of misspecified MLE (misMLE) in the censored sample cases. Since the ALT studies involve extrapolations, the mis-specification of the regression model could be more important than

the mis-specification of the lifetime distribution. For this reason, we focus our study on the mis-specified regression model.

Suppose that the main interest is to estimate the 10% lifetime at the normal-use condition. Our goal is to compare the percentile regression against the mis-specified model, mean (and variance) regression. When the lifetime distribution is in a location-scale family, the percentile is a linear function of location and scale parameters. In the case of a constant scale, the percentile is a linear function of the location parameter, and the mean and percentile regressions will lead to the same result. When the lifetime distribution is not in the location-scale family, the separation between the mean and percentile regression can be significant. For example, consider the $Gamma(\theta, \kappa)$ distribution, where $\theta > 0$ is a scale parameter and $\kappa > 0$ is a shape parameter. The q-th percentile is $\xi_q = \theta\Gamma_I^{-1}(q;\kappa)$, where Γ_I is the incomplete Gamma function (Meeker & Escobar, 1998, page 99).

When the parametric model is mis-specified, the estimator obtained by maximizing the log-likelihood is no longer asymptotically optimal. Under proper regularity conditions, White (1982) showed that the misMLE will converge to a well defined limit of θ^*, which minimizes the Kullback-Leibler Information Criterion, $\mathrm{E}_g(\log[g(\theta, \boldsymbol{t})/f(\theta;\boldsymbol{t})])$, where f is the mis-specified pdf, g is the true pdf, and the expectations are taken with respect to the true distribution with the parameter θ. The asymptotic variance for the misMLE can be evaluated as

$$\Sigma(\theta^*) = A(\theta^*)^{-1}B(\theta^*)A(\theta^*)^{-1}, \tag{72}$$

where

$$\begin{aligned} A(\theta) &= \mathrm{E}_g(\partial^2 \log f(\boldsymbol{t},\theta)/\partial\theta_i\partial\theta_j), \\ B(\theta) &= \mathrm{E}_g[\partial \log f(\boldsymbol{t},\theta)/\partial\theta_i \cdot \partial \log f(\boldsymbol{t},\theta)\partial\theta_j]. \end{aligned}$$

Consider an ALT experiment with three levels, where low, middle, and high stress levels are re-scaled as 0.5, 0.75 and 1, respectively, such that the normal-use stress level $x_D = 0$. Replicated samples are allocated at each stress levels according to a $4:2:1$ proportion (Meeker and Hahn, 1985) of their sample sizes. We will consider the log-normal and Gamma distributions, representing location-scale and non-location-scale families, respectively.

Experiment #1 - Log-normal Distribution: Assume that the location and scale parameters in $LN(\mu, \sigma)$ can be modeled using linear function of stress levels x as $\mu(x) = 1 - x$ and $\sigma(x) = 2 - x$. Then, the $100q\%$ lifetime in log-scale is $\xi_q(x) = \mu(x) + z_q\sigma(x) = (1 + 2z_q) + (z_q - 1)x = x'\beta$, where z_q is the q-quantile of the standard normal distribution. Four regression cases and two censoring are explored.

1. $\mu(x) = x'\alpha$, $\sigma(x) =$ constant
2. $\mu(x) = x'\alpha$, $\log \sigma(x) = x'\gamma$
3. $\mu(x) = x'\alpha$, $\sigma(x) = x'\gamma$ (true parametric model)
4. $\xi_q(x) = x'\beta$, using the AMELE method to estimate β

The first three models are for the parametric MLE and the last one is for the AMELE. Since the percentile here is a function of both location and scale parameters, a mis-specification of the scale (or mean) model does not destroy the percentile's linear relationship with the stress variable. The first case studies the impact of a mis-specified variance model. The second case is similar to the first. However, when the variance model is nonlinear, the percentile regression becomes nonlinear.

Table 4, summarizing the proportions of failure data for the simulated data, shows that the MSE in the misMLE Cases 1(a-b) is more than 74% larger than the MSE in AMELE's in all cases studied. The bias in the MisMLE makes the MSE significantly larger than the AMELE's. Since the log-linear variance model used in Case 1(b) can approximate the true linear model better than the constant variance model, Case 1(b) has smaller bias. Compared with the correct parametric MLE, the AMELE's asymptotic variance is about 28% larger. When the model is mis-specified, the asymptotic variance could be seriously under- or -over estimated. For example, the variance in Case 1(b) is about 34% larger than the variance in the AMELE (Case 1(d)).

Experiment #2 - Gamma Distribution: Consider a non-location-scale $Gamma(\theta, \kappa)$ family of distributions, where $\theta > 0$ is a scale parameter and $\kappa > 0$ is a shape parameter.

Table 10: Simulation Results of Experiment #1

Proportion Failing	n=301	Case 1(a)	Case 1(b)	Case 1(c)	Case 1(d)
(0.70, 0.80, 0.90)	bias^2	0.557	0.102	0.00	0.00
	Var($\hat{\xi}_q$)	0.089	0.335	0.158	0.221
(0.44, 0.50, 0.60)	bias^2	0.334	0.143	0.00	0.00
	Var($\hat{\xi}_q$)	0.078	0.369	0.163	0.221

Table 11: Simulation Results of Experiment #2

Proportion Failing	n=301	Case 2(a)	Case 2(b)	Case 2(c)
(0.73, 0.82, 0.90)	bias^2	1.22	8.30	0.000
	Var($\hat{\xi}_q$)	0.38	3.77	0.25

Assume a linear percentile regression, $\xi_q(x_j) = \beta_0 + \beta_1 x_j = \theta_j \Gamma_I^{-1}(q, \kappa_j)$. For simplicity, we fix the $\kappa_j = \kappa$ for all stress levels, and choose $\kappa = 2$ indicating an increasing hazard rate function. Let $\beta_0 = 2$, $\beta_1 = -1$.

Although the Gamma distribution doesn't belong to the location-scale family, the following parametrization is often recommended for numerical stability: $\mu = \log(\theta)$, $\sigma = 1/\sqrt{\kappa}$. Thus, the scale parameters $\theta_L, \theta_M, \theta_H$ are $2.8205, 2.3505, 1.8804$, respectively. Since $\theta = exp(\mu)$ and the true model has a linear regression structure on percentile $\xi_q(x_j) = exp(\mu)\Gamma_I^{-1}(q, \kappa_j)$, $\theta = exp(\mu)$ is a linear function of stress when $\kappa_j = \kappa$. Three models are investigated below; the first two give simple, alternative structures of the regression function of μ.

1. $\mu = x'\alpha$, $\sigma =$ constant.

2. $\log(\mu) = x'\alpha$ (or $\mu = exp(x'\alpha)$), $\sigma =$ constant.

3. $\xi_q(x) = x'\beta$, using AMELE method to estimate β.

Table 11 shows that the misspecification leads to an even larger asymptotic bias and variance compared to results in Experiment #1. Both cases in 2(a-b) have larger variance than the AMELE.

5.5.2 PCB Example

In Section 2, Figure 24 showed that Weibull distribution is not appropriate for modeling the PCB data from Meeker and LuValle (1995). We consider it further, along with the AMELE method, because it serves as benchmark for testing the AMELE. Following Meeker and LuValle's Weibull mean regression model,

$$F_{T_j}(t;\beta_0,\beta_1,\sigma) = \Phi_{EV}(Y_j), \quad Y_j = (log(t) - \mu(X_j))/\sigma,$$

where

$$\mu(X_j) = \beta_0 + \beta_1 \text{logit}(X_j), \quad X_j = \text{RH}/100, \quad \text{logit}(p) = \log[p/(1-p)], \tag{73}$$

and Φ_{EV} is the cdf of the standard extreme value distribution. In this model, σ is the same at all levels, and the logit-transformation can be justified physically (Meeker and LuValle, 1995). The parametric MLE estimates for model parameters are calculated as $\hat{\beta}_0 = 9.10$, $\hat{\beta}_1 = -3.78$, $\hat{\sigma} = 0.93$, respectively.

Figure 26 gives profile likelihood plots for each of the three Weibull regression parameters. The horizontal lines on Figure 26 are drawn such that the their intersection with the profile likelihood provide approximate 95% confidence intervals (CIs) based on inverting the likelihood-ratio (LR) test. Using these plots, one can obtain the 95% LR-based CIs for β_0, β_1 and σ as $(8.82, 9.43)$,$(-4.17, -3.43)$ and $(0.83, 1.05)$, respectively.

The estimate of the pth quantile η_p of $Y = \log(T)$ is $\hat{\eta}_p(x) = \hat{\beta}_0 + \hat{\beta}_1 + w_p\hat{\sigma}$, where $w_p = \log[-\log(1-p)]$. Confidence intervals for $\xi_p(x)$ can be obtained by using the large-sample normal approximation with the asymptotic variance calculated from the Fisher information matrix (Lawless 1982, page 305). Under the normal-use condition (RH=10%), the point estimate and CIs for the 5th percentile $\eta_{0.05}$ are calculated as 14.64 and $(13.54, 15.70)$, respectively.

Figure 27 compares the confidence intervals for the percentile regression coefficients ξ_p and β_1 using AMELE method. Using the delta method, the corresponding point estimate and CI of the 5th percentile lifetime at the normal-use condition are 12.30 and $(11.83, 12.78)$, respectively. Note that the width of this CI is only about 44% of the width for the CI calculated using the Weibull regression model. Back-transform the estimate to the original time

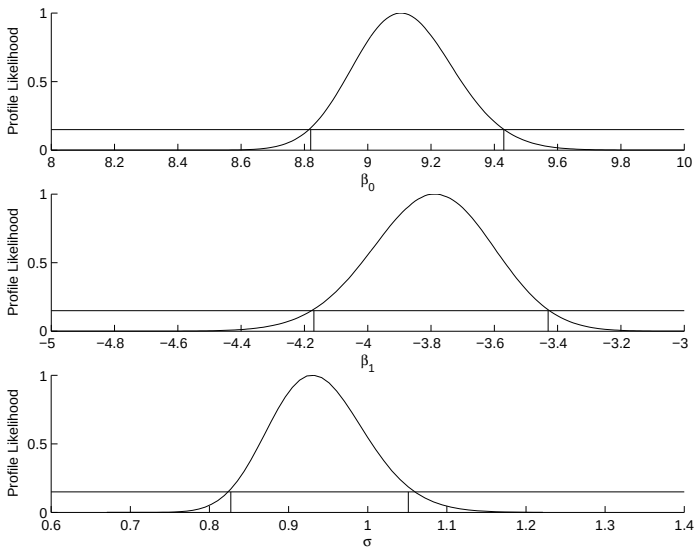

Figure 26: Profile Likelihoods of β_0, β_1 and σ Using the Weibull Regression Model

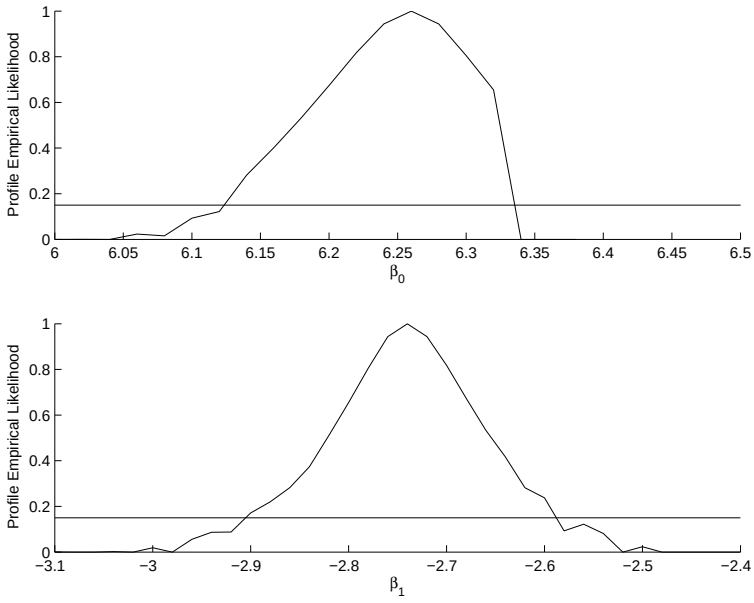

Figure 27: Profile Empirical Likelihoods for β_1 and $\xi_{0.05}$ Using the AMELE Method

scale, the 5th percentile lifetime is predicted as 25 years. Recall that based on the physics-based kinetics model given in Meeker and LuValle (1995), the proportion of product failing is less than 1% under the normal-use condition, and their prediction of the 5th percentile is infinity. Comparing this result against with the AMELE prediction, the AMELE is more conservative and arguably easier to interpret.

To validate the estimating equations pertaining to the log-5th-percentile's linear regression relationship with the stress levels, i.e., $\mathrm{E}(G_j(T_j; \boldsymbol{\theta}, X_j)) = 0$, the AMELE-based LR-test can be conducted according to Corollary 3.5. The LR statistic is $\Lambda_3 = 2.26$ and the p-value is 0.13. In this example, the percentile regression model seems plausible.

Next, we explore the difference in predicting the survival functions. Specifically, we examine the survival function of the failure time at the normal-use condition under different distribution assumptions. The data exploration analysis in Figure 2 of Section 2 shows that the 5th percentile regression and quantile-range regression provide possible adjustments for location and scale of the lifetime distributions at three stress levels. Consider the following two cases for this comparison.

1. Case (i) − After adjusting the 5th percentiles, lifetime distributions are the same.

2. Case (ii) − After adjusting the 5th percentiles and re-scaling with the quantile-range, lifetime distributions are the same.

Both cases can be justified by applying the nonparametric two-sample Wilcoxon-test to the adjusted-data at the higher stress levels. The AMELE in Case (i) estimates the 5th percentile regression parameters as $(\beta_0, \beta_1) = (6.2931, -2.6378)$. Correspondingly, the AMELE in Case (ii) leads to $(6.3060, -2.6753)$. Their prediction of the 5th percentile lifetime are 20 and 22 years for Case (i) and (ii), respectively. Note that with the adjustment from the scale, the lifetime distribution in Case (ii) should be much more spread out than the one in Case (i). This shows in the estimates of the survival function plotted in Figure (28). Figure (29) provides the point-wise confidence intervals for the survival function in Case (ii). Because there are more censored observations in the right tail, those intervals are larger than the ones in the left tail.

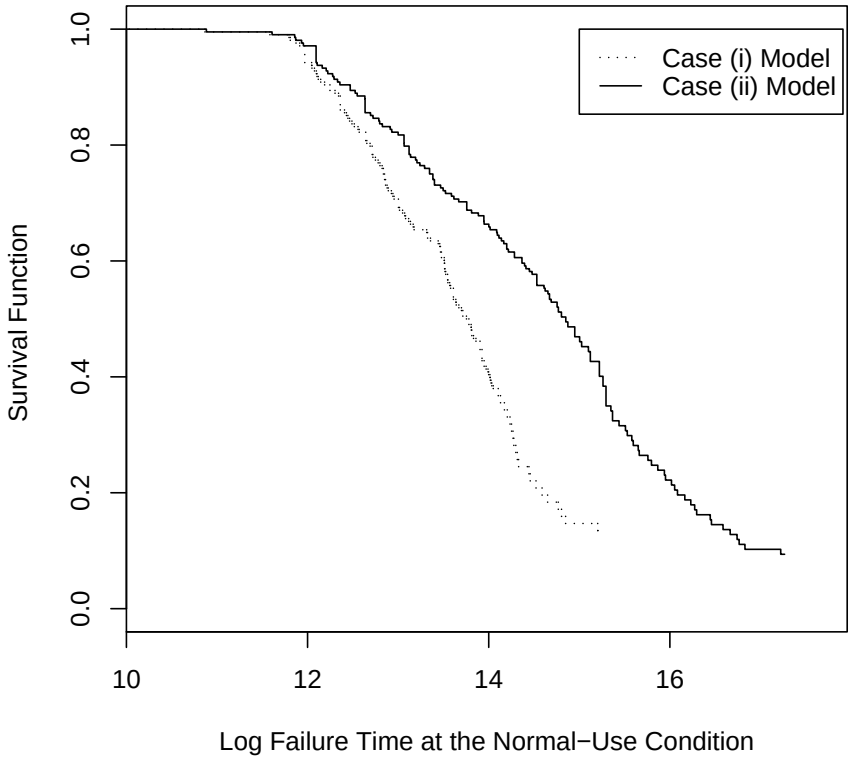

Figure 28: AMELE of Survival Functions Under Different Assumptions.

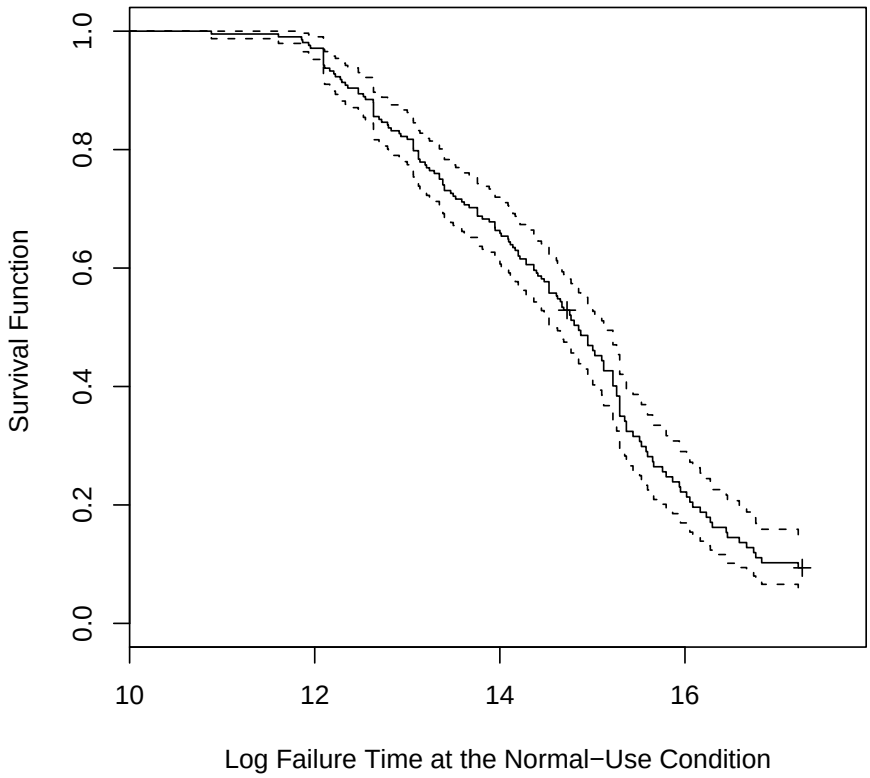

Figure 29: AMELE of Survival Function of the Failure Time

5.6 Concluding Remarks

The AMELE method provides reasonable estimates in a real-life example useful in reliability improvement comparisons. The proposed data-exploration based percentile and quantile-range regressions are effective in overcoming the difficulty of observing mean lifetime in the heavily censored data case for constructing commonly used mean and variance regression models in ALT studies. The LR-based test provides a revenue for validating the model formally. The asymptotic bias and variance studies show that the AMELE is reasonably competitive against the parametric MLE method when the model is correctly specified, and performs much better than the MLE when the model is incorrectly specified. Based on the properties derived in this article, the AMELE method should be a strong candidate for handling challenging data modeling and statistical inference problems.

5.7 Appendix:

Appendix A: Relationship of AMELE Estimator and Kaplan-Meier Estimator

From equations (54) - (56), it is clear that $\boldsymbol{\lambda}_j$ play an important role in the AMELE. If we set $\boldsymbol{\lambda}_j = 0$, i.e., no constraint is imposed on the empirical likelihood, the SMLE reduces to the NPMLE. Then, the estimate of the survival function (65) reduces to

$$\tilde{S}_{T_j,0}(t) = \sum_{t_{ij}>t} \frac{1}{n_j(1-a_{ij}(0))} \tag{74}$$

where "0" is placed in the subscript to emphasize $\boldsymbol{\lambda}_j = 0$, and

$$a_{ij}(0) = \frac{1}{n}\sum_{v=1}^{i} \frac{c_{vj}}{\sum_{l=v}^{k_j+1} P_{lj,0}},\ i = 2, \cdots, k_j+1, \quad a_{n1}(0) = 0, \tag{75}$$

$$P_{ij,0} = \frac{1}{n_j(1-a_{ij}(0))}, \quad i = 1, 2, \ldots, k_j,\ P_{(k_j+1)j,0} = 1 - \sum_{i=1}^{k_j} P_{ij,0}.$$

When additional information about the lifetime distribution regarding to parameter $\boldsymbol{\theta}$ can be provided by the estimating equations, the AMELE will incorporate it through $\boldsymbol{\theta}$ via $\boldsymbol{\lambda}_j$ and $\boldsymbol{G}_{jh}(t, \boldsymbol{\theta})$.

The following show that $\tilde{S}_{T_j,0}(t)$ is the well-known Kaplan-Meier estimate (Kaplan and Meier, 1958). Denote by $W_{ij} = 1 - \sum_{v=1}^{i} P_{vj}$, $\beta_{ij} = W_{ij}/W_{(i-1)j}$, $i = 1, 2, \ldots, k_j+1$, $j =$

$1, 2 \ldots, m$, so that $P_{ij} = W_{ij} - W_{(i-1)j} = (1-\beta_{ij}) \prod_{v=1}^{i-1} \beta_{vj}$, and $1 - \sum_{v=1}^{i} P_{vj} = \prod_{v=1}^{i} \beta_{vj}$. With this notation, it follows from (74) and (75) that for $i = 2, 3 \ldots, k_j, \quad j = 1, 2 \ldots, m$,

$$\begin{aligned} 1-\beta_{ij} &= \frac{1}{\prod_{v=1}^{i-1} \beta_{vj} (n_j - \sum_{u=1}^{i} c_{uj} / \prod_{s=1}^{v} \beta_{sj})}, \\ 1-\beta_{1j} &= P_{1j} = \frac{1}{n - c_{1j}}. \end{aligned} \tag{76}$$

Note that, from (76), β_{ij} can be expressed in terms of $\beta_{1j}, \beta_{2j}, \ldots, \beta_{(i-1)j}$, and $\beta_{1j} = (n-1-c_{1j})/(n-c_{1j})$. After further simplification,

$$\beta_{ij} = \frac{n_j - \sum_{v=1}^{i} (c_{vj}+1)}{n_j - \sum_{v=1}^{i-1} (c_{vj}+1) - c_{ij}}, \quad 2, 3, \ldots, k_j, \quad j = 1, 2 \ldots, m.$$

Denote by $n_{ij} = n_j - \sum_{v=1}^{i-1} (c_{vj}+1) - c_{ij}$ the number of subjects at risk at time t_{ij} and let $n_{1j} = n_j$. Then β_{ij} can be written as $(n_{ij}-1)/n_{ij}$, and $\tilde{S}_{T_j,0}(t)$ can be expressed as a product limit estimator:

$$\tilde{S}_{T_j,0}(t) = \sum_{t_{ij} > t} P_{ij} = \prod_{t_{ij} \le t} \beta_{ij} = \prod_{t_{ij} \le t} \frac{(n_{ij}-1)}{n_{ij}} = \hat{S}_{T_j,KM}(t). \tag{77}$$

Consider $a_{ij}(0)$ defined in (75). It follows from (77) that

$$a_{ij}(0) = \sum_{v=1}^{i-1} \frac{c_{vj}}{n_j \hat{S}_{T_j,KM}(t_{vj})} = \sum_{v=1}^{i-1} \frac{c_{vj}}{n_j \prod_{l=1}^{v} \frac{(n_{lj}-1)}{n_{lj}}}, \quad i = 2, 3, \ldots, k+1.$$

With the same notation used before, let $W_{vj} = 1 - a_{vj,0}$, $\beta_{vj} = W_{vj}/W_{(v-1)j}$, $i = 2, 3, \ldots, k+1$. Then,

$$Q_{vj} = W_{vj} - W_{v-1,j} = \frac{c_{v-1,j}}{n_j \prod_{l=1}^{v-1} \frac{(n_{lj}-1)}{n_{lj}}} = (1-\beta_{vj}) \prod_{l=1}^{(v-1)j} \beta_{lj},$$

and $1 - a_{ij}(0) = 1 - \sum_{v=1}^{i-1} Q_{vj} = \prod_{v=1}^{i-1} \beta_{vj}$. It follows that

$$1-\beta_{ij} = \frac{c_{i-1,j}}{n_j \prod_{v=1}^{i-1} \frac{(n_{vj}-1)}{n_{vj}}} \prod_{v=1}^{i-1} \beta_{vj}, \quad i = 3, 4, \ldots, k+1, \quad \text{and} \quad 1-\beta_{2j} = \frac{c_{1j}}{n_{1j} - c_{1j}},$$

where $n_{ij} = n_j - \sum_{v=1}^{i-1} (c_{vj}+1)$, for $i > 1$, the number of subjects at risk at time point t_{ij} and $n_{1j} = n_j$, then the β_{ij} can be written as $(n_{ij} - c_{ij})/n_{ij}$. Therefore,

$$1 - a_{ij,0} = \sum_{v=i}^{k+1} Q_{vj} = \prod_{v=1}^{i-1} \beta_{vj} = \prod_{v=1}^{i-1} \frac{(n_{ij} - c_{vj})}{n_{ij}} = \hat{S}_{C_j,KM}(t_{ij}), \tag{78}$$

which is the Kaplan-Meier estimate of the survival function $S_{C_j}(t)$ at time t_{ij}. (77) and (78) will play critical roles in deriving the asymptotic distribution of AMELE. □

Appendix B: Properties of Gaussian Processes

Lemma B.1 *Let $W(t)$ be a Gaussian process satisfying (1) $W(0) = 0$ and $E(W(t)) = 0$ for any t; (2) for any (s,t),*

$$Cov(W(s), W(t)) = R(t)R(s) \int_0^{\min(s,t)} v(x)dx.$$

Let $G_h(t) = (g_{1h}(t), \dots g_{rh}(t))^\top$ and denote

$$\Psi_G = \int_0^\infty W(t)dG_h(t) = \left[\int_0^\infty W(t)dg_{1h}(t), \dots, \int_0^\infty W(t)dg_{rh}(t)\right]^\top.$$

Suppose that $G_h(t)$ is differentiable, and $G_h(t) \to G(t)$ as $h \to 0$. Suppose $R(t) \to 0$ as $t \to \infty$. Then, Ψ_G is distributed $N_m(0, \mathbf{\Sigma}_G)$, where

$$\mathbf{\Sigma}_G = \int_0^\infty v(s) \left[\int_s^\infty (G(s) - G(x))dR(x)\right] \left[\int_s^\infty (G(s) - G(x))dR(x)\right]^\top ds.$$

Proof: We show the normality first. Let $0 = x_0 < x_1 < \cdots < x_n$, and $\Psi_n = \sum_{i=1}^n W(x_i)(G_h(x_i) - G_h(x_{i-1}))$. Then Ψ_n is a series of normal random variables. Suppose that, as $n \to \infty$, $\max_{1\le i\le n} |x_i - x_{i-1}| \to 0$ and $x_n \to \infty$. Then we know that $\Psi_n \xrightarrow{p} \Psi_G$. It follows from the normality of every Ψ_n that Ψ_G is normal distributed as well, and $\mathrm{E}(\Psi_G) = \int_0^\infty \mathrm{E}(W(x))dG_h(x) = 0$.

Next, we calculate the covariance of Ψ_G. Note that

$$\begin{aligned}
\mathbf{\Sigma}_G &= \mathrm{Cov}\left(\int_0^\infty W(x)dG_h(x), \int_0^\infty W(x)dG_h(x)^\top\right) \\
&= \mathrm{E}\left[\int_s^\infty W(t)dG_h(t)\right] \left[\int_s^\infty W(s)dG_h(s)\right]^\top \\
&= \int_0^\infty \int_0^\infty \mathrm{E}(W(t)W(s))dG_h(s)dG_h(t)^\top \\
&= \int_0^\infty \int_0^\infty \left(R(t)R(s) \int_0^{\min(s,t)} v(x)dx\right) dG_h(s)dG_h(t)^\top \\
&= \int\int_{t\le s} \left(R(t)R(s) \int_0^t v(x)dx\right) dG_h(s)dG_h(t)^\top + \\
&\quad \int\int_{t>s} \left(R(t)R(s) \int_0^s v(x)dx\right) dG_h(s)dG_h(t)^\top \\
&= \int_0^\infty v(x)dx \left(\int_x^\infty R(t) \int_t^\infty R(s)dG_h(s)dG_h(t)^\top\right) + \\
&\quad \int_0^\infty v(x)dx \left(\int_x^\infty R(s) \int_x^t R(t)dG_h(t)dG_h(s)^\top\right) \\
&= \int_0^\infty v(x)dx \left[\int_x^\infty R(t)dG_h(t)\right] \left[\int_x^\infty R(t)dG_h(t)\right]^\top.
\end{aligned}$$

Integrating by parts, we know that

$$\begin{aligned}\int_x^\infty R(t)dG_h(t) &= -R(x)G_h(x) - \int_x^\infty G_h(t)dR(t) = \int_x^\infty G_h(x)dR(t) - \int_x^\infty G_h(t)dR(t) \\ &= \int_x^\infty (G_h(x) - G_h(t)dR(t) = \int_x^\infty (G(x) - G(t)dR(t).\end{aligned}$$

It follows that

$$\begin{aligned}\mathbf{\Sigma}_G &= \mathrm{Cov}\left(\int_0^\infty W(x)dG_h(x), \int_0^\infty W(x)dG_h(x)^\top\right) \\ &= \int_0^\infty v(s)\left[\int_s^\infty (G(s) - G(x))dR(x)\right]\left[\int_s^\infty (G(s) - G(x))dR(x)\right]^\top ds.\end{aligned}$$

The proof is thus completed. □

Denote $W_n(t) = \sqrt{n}(\hat{S}_T(t) - S_T(t))$, where $\hat{S}_T(t)$ is the Kaplan-Meier estimator of $S_T(t)$. We know $W_n(t)$ converges to a Gaussian process $W(t)$ satisfying $W(0) = 0$ and $\mathrm{E}(W(t)) = 0$ with covariance function

$$\mathrm{Cov}(W(s), W(t)) = S_T(t)S_T(s)\int_0^{\min(t,s)} \frac{dF_T(x)}{S_T^2(x)S_C(x)}.$$

Replacing $R(t)$ and $v(x)dx$ in (79) by $S_T(t)$ and $S_T(x)^{-2}S_C(x)^{-1}dF_T(x)$, respectively, we obtain the following corollary.

Lemma B.2 *Let* $\boldsymbol{W}_n = \int_0^\infty \sqrt{n}(\hat{S}_T(x) - S_T(x))dG_h(x)$. $G_h(x) \to G(x)$ *as* $h \to 0$ *for almost every* x. *Then* $\boldsymbol{W}_n$ *has the asymptotic normal distribution* $N(0, \mathbf{\Sigma}_G)$, *with covariance matrix*

$$\mathbf{\Sigma}_G = \int_0^\infty \left[\int_x^\infty (G(x) - G(t))dS_T(t)\right]\left[\int_x^\infty (G(x) - G(t))dS_T(t)\right]^\top \frac{dF_T(x)}{S_T^2(x)S_C(x)}.$$

□

Appendix C: Proof of Lemmas and Theorems

In the following proofs for Lemmas, subscript j is dropped for notational simplicity.

Proof of Lemma 1: First, we evaluate the $\mathrm{E}G_h(T,\boldsymbol{\theta}_0)$ and $\mathrm{E}\dfrac{\partial G_h(T,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}$. Let F_T denote the cumulative distribution function for T. According to the percentile regression in (51), the qth percentile of $T - x'\boldsymbol{\theta}_0$ is 0. If we denote F_0 to be the distribution function for $T - x'\boldsymbol{\theta}_0$, then $F_0(0) = q$. By changing the order of the integral variables,

$$\begin{aligned}\mathrm{E}G_h(T,\boldsymbol{\theta}_0) &= \int K(u)F_T(x'\boldsymbol{\theta}_0 - hu)du - \int q du \\ &= \int K(u)\left[F_0(-hu) - F_0(0)\right]du.\end{aligned}$$

Then, apply a Taylor expansion of F_0 at 0, $F_0(-hu) = F_0(0)+f_0(0)(-hu)+\cdots+f_0^{r-1}(0)(-hu)^r+ o(h^r)$. Due to the property of $K(u)$ in (52), $\mathrm{E}G_h(T,\boldsymbol{\theta}_0) = (-h)^r(r!)^{-1}\kappa f_0^{r-1}(0) + o(h^r) = o(n^{-1/2})$. Similarly, we can show that $\mathrm{E}\dfrac{\partial G_h(T,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} = \boldsymbol{C}(\boldsymbol{\theta}_0) + o(1) = f(0)x' + o(1)$, which proves part (c).

In the small neighborhood of $\boldsymbol{\theta}_0$,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{k+1} Z_{i,h}(\boldsymbol{\theta}) &= \int G_h(t,\boldsymbol{\theta})d\hat{F}_{T,KM}(t) \\ &= \int G_h(t,\boldsymbol{\theta}_0)d\hat{F}_{T,KM}(t) + \int \frac{\partial G_h(t,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)d\hat{F}_{T,KM}(t) + o(n^{-1/2}) \\ &= \int G_h(t,\boldsymbol{\theta}_0)d\hat{F}_{T,KM}(t) - \int G_h(t,\boldsymbol{\theta}_0)dF_T(t) + O(n^{-1/2}) \\ &= \int (\hat{S}_{T,KM}(t) - S_T(t))dG_n(t,\boldsymbol{\theta}_0) + O(n^{-1/2}),\end{aligned}$$

where $\sqrt{n}(\hat{S}_{T,KM}(t) - S_T(t))$ converges to a normal process $W(t)$ defined in Lemma $B.1$. Thus, part (a) is proved.

As shown in Appendix A, $1 - a_i(0) = \hat{S}_{C,KM}(t_i)$, we can write

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{k+1} Z_{ih}(\boldsymbol{\theta})Z_{ih}(\boldsymbol{\theta})^\top &= \int \frac{G_h(t,\boldsymbol{\theta})G_h(t,\boldsymbol{\theta})^\top}{\hat{S}_{C,KM}(t)}d\hat{F}_{T,KM}(t) \qquad (79) \\ &\to \int \frac{G(t,\boldsymbol{\theta})G(t,\boldsymbol{\theta})^\top}{S_C(t)}dF_T(t) = \boldsymbol{A}(\boldsymbol{\theta}_0). \qquad (80)\end{aligned}$$

It follows from the uniform consistency of Kaplan-Meier estimate and the bounded derivatives of $G_h(t,\boldsymbol{\theta})$, part (b) can be proved. □

Proof of Lemma 2: $\boldsymbol{\lambda}$ is solved from the implicit function $\sum_{i=1}^{k+1} \frac{Z_{i,h}(\boldsymbol{\theta})}{1+\boldsymbol{\lambda}^\top Z_{i,h}(\boldsymbol{\theta})} = 0$. Since we have $\frac{1}{n}\sum_{i=1}^{k+1} Z_{ih}(\boldsymbol{\theta})Z_{ih}(\boldsymbol{\theta})\top < \infty$, it is easy to verify that $max_i||Z_{ih}(\boldsymbol{\theta}) = o(n^{1/2})$. Following the steps used in Owen (1990), we can establish that

$$\frac{||\boldsymbol{\lambda}||}{1+||\boldsymbol{\lambda}||\ \max_i||Z_{ih}(\boldsymbol{\theta})||} = O_p(n^{-1/2}),$$

which implies that $||\boldsymbol{\lambda}|| = O_p(n^{-1/2})$.

Proof of Theorem 1:

According to (66), $\boldsymbol{\lambda}_j = -A_j(\boldsymbol{\theta}_0)^{-1}\left[\int_0^\infty G_{jh}(t,\boldsymbol{\theta})d\hat{F}_{T_j,KM}(t)\right] + o_p(n-1/2)$. Because

$$\mathrm{E}[G_{jh}(T_j,\boldsymbol{\theta})] = \int_0^\infty G_{jh}(t,\boldsymbol{\theta})dF_{T_j}(t) = 0$$

as $h \to 0$, we know

$$\int_0^\infty G_{jh}(t,\boldsymbol{\theta})d\hat{F}_{T_j,KM}(t) = \int_0^\infty G_{jh}(t,\boldsymbol{\theta})d(\hat{F}_{T_j,KM}(t)-F_{T_j}(t)) = -\int_0^\infty G_j(t,\boldsymbol{\theta})d(\hat{S}_{T_j,KM}(t)-S_{T_j}(t)).$$

Using integration by parts, it follows that

$$\int_0^\infty G_{jh}(t,\boldsymbol{\theta})d\hat{F}_{T_j,KM}(t) = \int_0^\infty (\hat{S}_{T_j,KM}(t) - S_{T_j}(t))dG_{jh}(t,\boldsymbol{\theta}). \tag{81}$$

According to Breslow and Crowley (1974, Theorem 5), $\sqrt{n}(\hat{S}_{T_j,KM}(t) - S_{T_j}(t))$ converges to a Gaussian process $W_j(t)$, with $\mathrm{E}(W_j(t)) = 0$ and

$$\mathrm{Cov}(W_j(s),W_j(t)) = S_{T_j}(s)S_{T_j}(t)\int_0^{\min(t,s)} \frac{dF_{T_j}(x)}{(S_{T_j}(x)^2 S_{C_j}(x))}. \tag{82}$$

It follows from (81) that (under condition **R3**) as $n_j \to \infty$,

$$\sqrt{n_j}\int_0^\infty G_{jh}(t,\boldsymbol{\theta})d\hat{F}_{T_j,KM}(t) \xrightarrow{p} \int_0^\infty W_j(t)dG_{jh}(t,\boldsymbol{\theta}).$$

Using Gaussian process properties (see Appendix Lemma B.2 for details), we can obtain that $\int_0^\infty W_j(t)\ dG_{jh}(t,\boldsymbol{\theta})$ is normal with mean zero and covariance matrix $\boldsymbol{\Sigma}_{G_j}(\boldsymbol{\theta})$ defined in (68) (see Appendix Corollary B.2). Thus, $\sqrt{n_j}\boldsymbol{\lambda}_j$ is asymptotic normal with mean zero and covariance matrix

$$\boldsymbol{\Sigma}_{\lambda_j}(\boldsymbol{\theta}) = \boldsymbol{A}_j(\boldsymbol{\theta})^{-1}\boldsymbol{\Sigma}_{G_j}(\boldsymbol{\theta})\boldsymbol{A}_j(\boldsymbol{\theta})^{-1},$$

where $\boldsymbol{A}_j(\boldsymbol{\theta})$ is given by (79). This completes the proof. □

Proof of Theorem 4:

Again, consider $\boldsymbol{\theta}$ in a small ball $\{\boldsymbol{\theta} : ||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| \leq n^{-1/2}$. Because $\boldsymbol{\lambda}_j(\boldsymbol{\theta}) = O_p(n^{-1/2})$, we make the Taylor expansion of $\tilde{S}_{T_j}(t)$ at $\boldsymbol{\lambda}_j = 0$ following the similar procedure in Qin and Lawless (1994), which results in

$$\begin{aligned}\tilde{S}_{T_j}(t) &= \sum_{t_{ij}>t}\left(\frac{1}{n(1-a_{ij}(0))} + \frac{G_{jh}^{\top}(t_{ij},\boldsymbol{\theta})\boldsymbol{\lambda}_j(\boldsymbol{\theta})}{n(1-a_{ij}(0))^2} + o_p(n^{-1/2})\right) \\ &= \hat{S}_{T_j,KM}(t) + \int_t^{\infty}\frac{G_{jh}^{\top}(x,\boldsymbol{\theta})d\hat{F}_{T_j,KM}(x)}{\hat{S}_{C_j,KM}(x))}\boldsymbol{\lambda}_j(\boldsymbol{\theta}) + o_p(n^{-1/2}),\end{aligned}$$

where $\hat{F}_{T_j,KM}(t) = 1 - \hat{S}_{nT_j,KM}(t)$, and $\hat{S}_{T_j,KM}(t)$ and $\hat{S}_{C_j,KM}(t)$ are the Kaplan-Meier estimates of $S_{T_j}(t)$ and $S_{C_j}(t)$. By replacing $\boldsymbol{\lambda}_j(\boldsymbol{\theta})$ with $-\boldsymbol{A}_j^{-1}\int_0^{\infty} G_{jh}(t,\boldsymbol{\theta})d\hat{F}_{T_j,KM}(t)$,

$$\tilde{S}_{T_j}(t) = \hat{S}_{T_j,KM}(t) + \left(\int_t^{\infty}\frac{G_{jh}^{\top}(x,\boldsymbol{\theta})d\hat{F}_{T_j,KM}(x)}{\hat{S}_{C_j,KM}(x))}\right)\boldsymbol{A}_j^{-1}(\boldsymbol{\theta})\int_0^{\infty} G_{jh}(t,\boldsymbol{\theta})d\hat{F}_{T_j,KM}(t) + o_p(n_j^{-1/2}),$$

where

$$\boldsymbol{A}_j(\boldsymbol{\theta}) = \int_0^{\infty}\frac{G_{jh}(t,\boldsymbol{\theta})G_{jh}^{\top}(t,\boldsymbol{\theta})}{\hat{S}_{C_j,KM}(t)}d\hat{F}_{T_j,KM}(t).$$

It follows that

$$\begin{aligned}\sqrt{n_j}(\tilde{S}_{T_j}(t) - S_{T_j}(t)) &= \sqrt{n_j}(\hat{S}_{T_j,KM}(t) - S_{T_j}(t)) + \\ &\quad \sqrt{n_j}\left(\int_t^{\infty}\frac{G_{jh}^{\top}(x,\boldsymbol{\theta})d\hat{F}_{T_j,KM}(x)}{\hat{S}_{C_j,KM}(x)}\right)\boldsymbol{A}_j^{-1}(\boldsymbol{\theta})\int_0^{\infty} G_{jh}(t,\boldsymbol{\theta})d\hat{F}_{T_j,KM}(t).\end{aligned}$$

Denote

$$\boldsymbol{\beta}_j(t,\boldsymbol{\theta}) = \lim\boldsymbol{A}_j^{-1}(\boldsymbol{\theta})\int_t^{\infty}\frac{G_{jh}(x,\boldsymbol{\theta})d\hat{F}_{T_j,KM}(x)}{\hat{S}_{C_j,KM}(x)} = \boldsymbol{A}_j^{-1}(\boldsymbol{\theta})\int_t^{\infty}\frac{G_j(x,\boldsymbol{\theta})dF_{T_j}(x)}{S_{C_j}(x)},$$

as $n_{min} \to \infty$. Then, we have

$$\sqrt{n_j}(\tilde{S}_{T_j}(t) - S_{T_j}(t)) = \sqrt{n_j}(\hat{S}_{T_j,KM}(t) - S_{T_j}(t)) + \sqrt{n_j}\boldsymbol{\beta}_j^{\top}(t,\boldsymbol{\theta})\int_0^{\infty} G_{jh}(x,\boldsymbol{\theta})d\hat{F}_{T_j,KM}(x).$$

The asymptotic normality follows from the fact that both $W_{n1j}(t) = \sqrt{n_j}(\hat{S}_{nT_j,KM}(t) - S_{T_j}(t))$ and $W_{n2j}(t) = \sqrt{n_j}\boldsymbol{\beta}_j^{\top}(t,\boldsymbol{\theta})\int_0^{\infty} G_{jh}(x,\boldsymbol{\theta})d\hat{F}_{T_j,KM}(x)$ are asymptotic normal, as $n_{min} \to \infty$.

Note that, according to Breslow and Crowley (1974, Theorem 5), $W_{n1j}(t)$ converges to a Gaussian process $W_{1j}(t)$, with $\mathrm{E}(W_{1j}(t)) = 0$,

$$\mathrm{Cov}(W_{1j}(s), W_{1j}(t)) = S_{T_j}(s)S_{T_j}(t)\int_0^{\min(t,s)}\frac{dF_T(x)}{S_{T_j}(x)^2 S_{C_j}(x)},$$

and

$$\begin{aligned}\lim_{n_j\to\infty} W_{n2j}(t) &= \lim_{n_j\to\infty} \sqrt{n_j}\boldsymbol{\beta}_j^\top(t,\boldsymbol{\theta}) \int_0^\infty G_{jh}(x,\boldsymbol{\theta}) d\hat{F}_{T_j,KM}(x) \\ &= \lim_{n_j\to\infty} \boldsymbol{\beta}_j^\top(t,\boldsymbol{\theta}) \int_0^\infty \sqrt{n_j}(\hat{S}_{nT_j,KM}(x) - S_{T_j}(x)) dG_{jh}(x,\boldsymbol{\theta})\end{aligned}$$

converges to a Gaussian process $W_{2j}(t)$, where

$$W_{2j}(t) = \boldsymbol{\beta}_j^\top(t,\boldsymbol{\theta}) \int_0^\infty W_{1j}(x) dG_{jh}(x,\boldsymbol{\theta}).$$

Therefore,

$$\sqrt{n_j}(\tilde{S}_{nT_j}(t) - S_{T_j}(t)) = W_{n1j}(t) + W_{n2j}(t) \xrightarrow{p} W_{1j}(t) + W_{2j}(t).$$

Note that $\mathrm{E}(W_{1j}(t)+W_{2j}(t)) = 0$, so that the asymptotic variance of $\sqrt{n_j}(\tilde{S}_{T_j}(t)-S_{T_j}(t))$ reduces to

$$\sigma_{\tilde{S}_j(t)} = \mathrm{Var}(W_{1j}(t)) + \mathrm{Var}(W_{2j}(t)) + 2\mathrm{Cov}(W_{1j}(t), W_{2j}(t)),$$

where

$$\begin{aligned}\mathrm{Cov}(W_{1j}(t), W_{2j}(t)) &= \mathrm{E}(W_{1j}(t)W_{2j}(t)) = \boldsymbol{\beta}_j^\top(t,\boldsymbol{\theta}) \int_0^\infty \mathrm{E}(W_{1j}(t)W_{1j}(x)) dG_{jh}(x,\boldsymbol{\theta}) \\ &= \boldsymbol{\beta}_j^\top(t,\boldsymbol{\theta}) \left(\int_0^t \mathrm{E}(W_{1j}(t)W_{1j}(x)) dG_{jh}(x,\boldsymbol{\theta}) + \int_t^\infty \mathrm{E}(W_{1j}(t)W_{1j}(x)) dG_{jh}(x,\boldsymbol{\theta}) \right) \\ &= \boldsymbol{\beta}_j^\top(t,\boldsymbol{\theta}) \left\{ \int_0^t \left(S_{T_j}(t)S_{T_j}(x) \right) \int_0^x \frac{dF_{T_j}(s)}{S_{T_j}(s)^2 S_{C_j}(s)} \right) dG_{jh}(x,\boldsymbol{\theta}) + \\ &\quad \int_t^\infty \left(S_{T_j}(t)S_{T_j}(x) \right) \int_0^t \frac{dF_{T_j}(s)}{S_{T_j}(s)^2 S_{C_j}(s)} \right) dG_{jh}(x,\boldsymbol{\theta}) \Big\} \\ &= S_{T_j}(t)\boldsymbol{\beta}_j^\top(t,\boldsymbol{\theta}) \left\{ \int_0^t \frac{dF_{T_j}(s)}{S_{T_j}(s)^2 S_{C_j}(s)} \int_s^t S_{T_j}(x)) dG_{jh}(x,\boldsymbol{\theta}) + \right. \\ &\quad \left. \int_0^t \frac{dF_{T_j}(s)}{S_{T_j}(s)^2 S_{C_j}(s)} \int_t^\infty S_{T_j}(x)) dG_{jh}(x,\boldsymbol{\theta}) \right\} \\ &= S_{T_j}(t) \int_0^t \left(\int_s^\infty \boldsymbol{\beta}_j^\top(t,\boldsymbol{\theta})(G_j(s,\boldsymbol{\theta}) - G_j(x,\boldsymbol{\theta}) dS_{T_j}(x) \right) \frac{dF_T(s)}{S_{T_j}(s)^2 S_{C_j}(s)}.\end{aligned}$$

Note that

$$\mathrm{Var}(W_{1j}(t)) = S_{T_j}^2(t) \int_0^t \frac{dF_{T_j}(x)}{S_{T_j}(x)^2 S_{C_j}(x)},$$

and it follows from Lemma B.1 that

$$\mathrm{Var}(W_{2j}(t)) = \int_0^\infty \left[\int_x^\infty \boldsymbol{\beta}_j^\top(t,\boldsymbol{\theta})(G_j(x,\boldsymbol{\theta}) - G_j(t,\boldsymbol{\theta})) dS_{T_j}(t) \right]^2 \frac{dF_{T_j}(x)}{S_{T_j}^2(x) S_{C_j}(x)}.$$

□

Appendix D:

Here, we benchmark our asymptotic results against some well-known results in the literature. Consider the one-sample case where $m = 1$. If θ is the population mean, the estimating function is $G(t, \theta) = t - \theta$. In this case, G is differentiable with regard to θ, so we don't require the smoothing parameter h. Because $\partial G/\partial\theta = 1$, $\boldsymbol{C}(\theta) = \boldsymbol{A}(\theta)^{-1}$ and $\boldsymbol{\Sigma}_l(\theta) = \boldsymbol{\Sigma}_\lambda(\theta)$, $\boldsymbol{\Sigma}_\theta(\theta)$ reduces to

$$\begin{aligned}\boldsymbol{\Sigma}_G(\theta) &= \int_0^\infty \left(\int_x^\infty (x-t))dS_T(t)\right)^2 \frac{dF_T(x)}{S_T^2(x)S_C(x)} \\ &= \int_0^\infty \left(\int_x^\infty (1-F_T(t))dt\right)^2 \frac{dF_T(x)}{S_T^2(x)S_C(x)},\end{aligned}$$

which is the same result as obtained by Breslow and Crowley (1974).

In the complete-sample case where $S_C(t) = 1$ and $G(\theta)$ is differentiable, (82) reduces to $\mathrm{Cov}(W(x), W(t)) = S_T(t)(1 - S_T(x))$. Then,

$$\begin{aligned}&\mathrm{Var}\left(\int_0^\infty W(t)dG(t,\boldsymbol{\theta})\right) \\ =\ & \int_0^\infty \left[\int_0^\infty \mathrm{E}(W(t)W(x))dG(x,\boldsymbol{\theta})\right] dG^\top(t,\boldsymbol{\theta}) \\ =\ & \int_0^\infty \left[\int_0^t S_T(t)(1-S_T(x))dG(x,\boldsymbol{\theta}) + \int_t^\infty S_T(x)(1-S_T(t))dG(x,\boldsymbol{\theta})\right] dG^\top(t,\boldsymbol{\theta}) \\ =\ & \int_0^\infty \left[\int_0^t S_T(t)dG(x,\boldsymbol{\theta}) + \int_t^\infty S_T(x)dG(x,\boldsymbol{\theta}) - \int_0^\infty S_T(x)S_T(t)dG(x,\boldsymbol{\theta})\right] dG^\top(t,\boldsymbol{\theta}) \\ =\ & \int_0^\infty \left[\int_0^t (S_T(t)-S_T(x))dG(x,\boldsymbol{\theta}) + (1-S_T(x))\int_0^\infty S_T(x)dG(x,\boldsymbol{\theta})\right] dG^\top(t,\boldsymbol{\theta}). \quad (83)\end{aligned}$$

Using integration by parts and the fact that

$$\int_0^\infty G(x,\boldsymbol{\theta}))dS_T(x) = -\int_0^\infty G(x,\boldsymbol{\theta}))dF_T(x) = \mathrm{E}[G(T,\boldsymbol{\theta})] = 0,$$

we have

$$\int_0^\infty S_T(x)dG(x\boldsymbol{\theta}) = -G(0,\boldsymbol{\theta}). \quad (84)$$

Again, using integration by parts, we know that

$$\int_0^t (S_T(t)-S_T(x))dG(x,\boldsymbol{\theta}) = (1-S_T(t))G(0,\boldsymbol{\theta}) - \int_0^t G(x,\boldsymbol{\theta})dF_T(x). \quad (85)$$

Plugging (84) and (85) into (83) and using integration by parts repeatedly, one can obtain that

$$\begin{aligned}\boldsymbol{\Sigma}_G(\boldsymbol{\theta}) &= \mathrm{Var}\left(\int_0^\infty W(t)dG(t,\boldsymbol{\theta})\right) = \int_0^\infty \left[\int_0^t G(x,\boldsymbol{\theta})dF_T(x)\right] dG^\top(t,\boldsymbol{\theta}) \\ &= \int_0^\infty G(t,\boldsymbol{\theta})G^\top(t,\boldsymbol{\theta})dF_T(t) = \mathrm{E}[G(T,\boldsymbol{\theta})G^\top(T,\boldsymbol{\theta})].\end{aligned}$$

Note that $\boldsymbol{A}(\boldsymbol{\theta}) = \mathrm{E}[G(T,\boldsymbol{\theta})G^\top(T,\boldsymbol{\theta})] = \boldsymbol{\Sigma}_G(\boldsymbol{\theta})$, and it follows that the asymptotic covariance matrix reduces to

$$\boldsymbol{\Sigma}_\theta = \left[\mathrm{E}\left(\frac{\partial G(T,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\mathrm{E}[G(T,\boldsymbol{\theta})G^\top(T,\boldsymbol{\theta})]^{-1}\mathrm{E}\left(\frac{\partial G(T,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^\top\right]^{-1},$$

which is the same result as obtained by Qin and Lawless (1994).

CHAPTER VI

FUTURE WORK BEYOND THESIS

There are several interesting topics that I would like to work on in the near future.

In the spatial smoothing procedure, the constrained optimization will lead to a smooth intensity function integrating to the constraints. This problem is similar to the smoothing splines problem which minimizes the penalized residual function. If theoretical proof can show that the spatial intensity function is a natural spline, then the computational complexity can be greatly reduce. In addition, the future research may include evaluation of model uncertainty and estimation uncertainty in affecting the optimal decisions by continuum approximation approaches.

The multi-level spatial model has been proposed to to characterize the hierarchy of aggregation in logistics planning. The combination of information was shown to provide more accurate predictions and better statistical inferences. It is interesting to see how can we construct new methods to provide better information fusion and integration at the multi-level decision framework.

Motivated by my intern experience in spare parts distribution at BellSouth Corporation, I feel it's important to develop a systematic method to integrate the reliability engineering and inventory management to achieve optimal system performance. The current inventory strategy of spare parts is based on the historical usage of products only, which often has very low demand for each product during a year. By proposing spare parts inventory strategy based on product reliability data, I expect that we can not only improve our estimation accuracy by aggregating historical demand for similar products, but also we can identify those products with unusual failure mechanism. This research work can be potentially developed into an integrated business solution package.

The inference on mixing models has been a challenging problem. With recent advancement in empirical likelihood, I would like to apply this new statistical inference methods

on estimating the mixing parameters. Due to the nice property of empirical likelihood, I expect to have some nice theoretical results on the asymptotic properties.

REFERENCES

[1] Abrahamsen, P. (1997), "A Review of Gaussian Random Fields and Correlation Functions", Report 917, Norvegian Computer Centre, Oslo.

[2] Bronnenberg, B.J., Mahajan, V. (2001), "Unobserved Retailer Behavior in Multimarket Data: Joint Spatial Dependence in Market Shares and Promotion Variables", *Marketing Science*, 20, 284-299.

[3] Bronnenberg, B.J., Sismeiro, C. (2002), "Using Multimarket Data to Predict Brand Performance in Markets for Which No or Poor Data Exist", *Journal of Marketing Research*, 39, 1-17.

[4] Daganzo, C. F. (1996), *Logisitcs Systems Analysis*, New York: Springer Verlag.

[5] Dandamudi, S., and Lu, J.-C. (2003), "Continuum Approximation of Logistics Operations for Supply-Chain Contract Decisions", paper presented in *2003 INFORMS Annual conference*, November, Atlanta, GA.

[6] Dasci, A., and Verter, V. (2001), "A Continuous Model for Production-Distribution System Design," *European Journal of Operational Research*, 129, 287-298.

[7] Delaney, R. V. (2004), *15th Annual State of Logistics Report*, Oak Brook, IL: Council of Logistics Management.

[8] Dyn, N., Wahba, G., and Wong, W.H. (1979), "Smooth pycnophylactic interpolation for geographical regions: comment," *Journal of the American Statistical Association*, 74, 530-535.

[9] Erera, A. L., and Daganzo, C. F. (2003), "A Dynamic Scheme for Stochastic Vehicle Routing," Technical Report in the School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA.

[10] Geoffrion, (1976), "The Purpose of Mathematical Programming is Insight, not Numbers", *Interface*, 7, 81-92.

[11] Gotway, C. A. and Young, L. J., (2002), "Combining Incompatible Spatial Data", *Journal of the American Statistical Association*, 97, 632-648.

[12] Lehmann, E. L., (1998), *Theory of Point Estimation*, New York: John Wiley.

[13] Langevin, A., Mbaraga, P., and Campbell, J. (1996), "Continuous Approximation Models in Freight Distribution: An Overview," *Transportation Research B*, 30, 163-188.

[14] Mangotra, D., and Lu, J.-C. (2005), "Coordinated Inventory Between National- and Regional Level DCs Based on Continnum Approximated Logistics Models," Technical Report, The School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA.

[15] Ouyang, Y. and Daganzo, C. F. (2003), "Discretization and validation of the continuum approximation scheme for terminal system design," Technical Report, Institute of Transportation Studies, University of California at Berkerley.

[16] Owen, A. B. (2001), *Empirical Likelihood*, Champman & Hall/CRC.

[17] Pawitan, Y. (2001), *In All Likelihood: Statistical Modeling and Inference Using Likelihood*, London: Oxford Science Publications.

[18] Qin, J. and Lawless, J. F. (1994), "Empirical Likelihood and General Estimating Equations," *Annal of Statistics*, 22, 300-325.

[19] Qin, J. (2000), "Combining parametric and empirical likelihoods," *Biometrika*, 87, 484-490.

[20] R Development Core Team (2004), "R: A language and environment for statistical computing," R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-00-3, URL http://www.R-project.org.

[21] U.S. Census Bureau: State & County Quick Facts, *http://quickfacts.census.gov/qfd/.*

[22] Wikle, C. K. and Berliner, L. M. (2005), "Combining Information Across Spatial Scales," *Technometrics*, 47, 80-91.

[23] E. Iritani, M. Dickerson, "Tallying Port Dispute's Costs", *Los Angeles Times*, 2002 Nov.

[24] B. M. Lewis, A. L. Erera, and C. C. White III, "An Inventory Control Model with Possible Border Disruptions", *Technical Report*, Department of Industrial & System Engineering, Georgia Institute of Technology, [Online], 2005. Available: http://www2.isye.gatech.edu/setra/reports/

[25] L. V. Snyder, M. S. Daskin, "Reliability models for facility location: The expected failure cost case," *Technical Report 04T-016*, 2004, Department of Industrial & System Engineering, Lehigh University.

[26] M. O. Ball, "Computing Network Reliability", *Operations Research*, vol 27, 1979 Jul & Aug, pp 823-838.

[27] B. Sanso, F. Soumis, "Communication and transportation network reliability using routing models", *IEEE Transactions on Reliability*, vol 40, 1991 Apr, pp 29-37.

[28] V. Shi, "Evaluating the performability of tactical communications network", *IEEE Transactions On Vehicular Technology*, vol 53, 2004 Jan, pp 253-260.

[29] K.K. Aggarwal, "A fast algorithm for the performance index of a telecommunication network", *IEEE Transactions on Reliability*, vol 27, 1988 Apr, pp 65-69.

[30] P.K. Varshney, A. R. Joshi, and P.L. Chang, "Reliability modeling and performance evaluation of variable link-capacity networks", *IEEE Transactions on Reliability*, vol 43, 1994 Sep, pp 378-382.

[31] T. Yokohira, M. Sugano, T. Nishida, and H. Miyahara, "Fault tolerant packet-switched network design and its sensitivity", *IEEE Transactions on Reliability*, vol 40, 1991 Oct, pp 452-460.

[32] S.C. Yang, J.A. Silvester, "Reconfigurable fault tolerant networks for fast packet switching", *IEEE Transactions on Reliability*, vol 40, 1991 Oct, pp 474-487.

[33] M. Mohorcic, M. Werner, A. Svigelj, and G. Kandus, "Adaptive routing for packet-oriented intersatellite link networks: performance in various traffic scenarios", *IEEE Transactions on Wireless Communications*, vol 1, 2002 Oct, pp 808-818.

[34] A. Chen, H. Yang, H. Lo, and W.H. Tang, "A capacity related reliability for transportation networks", *Journal of Advanced Transportation*, vol 33, 1999, pp 183-200.

[35] W. Q. Meeker, L. A. Escobar, *Statistical Methods for Reliability Data*, 1998, New York: John Wiley & Sons.

[36] M. Bundschuh, D. Klabjan, and D.L. Thurston, "Modeling robust and reliable supply chains," *Optimization Online* [Online], 2003 June. Available: http://www.optimization-online.org

[37] M.U. Thomas, "Supply chain reliability for contingency operations", *Proceedings Annual Reliability and Maintainability Symposium*, 2002 Jan, pp. 61-67.

[38] S. Dandamudi, J.C. Lu, "Competition driving logistics design with continuous approximation methods", Technical Report of the School of Industrial and Systems Engineering, Georgia Tech, [Online], 2004. Available: http://www.isye.gatech.edu/apps/research-papers/.

[39] C.F. Daganzo, A.L. Erera, "On planning and design of logistics systems for uncertain environments", *New Trends in Distribution Logistics*, vol 480 of *Lecture Notes in Economics and Mathematical Systems*, 1999 Jan, pp 3-21, Berlin: Springer-Verlag.

[40] W. R. Tobler, "Smooth pycnophylactic interpolation for geographical regions", *Journal of the American Statistical Association*, vol 74, 1979 Sep, pp 519-530.

[41] W. Kuo, M.J. Zuo *Optimal Reliability Modeling*, 2003, pp 157-164, New York: John Wiley & Sons.

[42] A.W. Bowman, A. Azzalini, "Computational aspects of nonparametric smoothing with illustrations from the sm library", *Computational Statistics and Data Analysis*, vol. 42, 2003 Apr, pp 545-560.

[43] N.A.C. Cressie, *Statistics for Spatial Data*, 1993, pp 651-656, New York: John Wiley & Sons.

[44] M.P. Wand, M.C. Jones, *Kernel Smoothing*, 1995, pp 60, Chapman & Hall.

[45] R Development Core Team (2005), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. [Online] Available: http://www.R-project.org

[46] Crow, L.H. (1974), *Reliability Analysis for Complex Repairable Systems, Reliability and Biometry*, eds. F. Proschan and R.J. Serfling, Philadelphia: SIAM, pp. 379-340.

[47] Duane, J. T. (1964), "Learning Curve Approach to Reliability", *IEEE Transactions on Aerospace*, 2, pp. 563-566.

[48] Engelhardt, M. and Bain, L.J. (1987), "Statistical Analysis of a Compound Power-Law Model for Repairable Systems", *IEEE Transaction on Reliability*, 36, pp. 392-396.

[49] Feng, Z. McCulloch, C.E. (1996), "Using Bootstrap likelihood Ratios in Finite Mixture Models", *Journal of the Royal Statistical Society, B*, 58, pp. 609-617.

[50] Finkelstein, J.M. (1976), "Confidence Bounds on the Parameters of the Weibull Process", *Technometrics*, 18, pp. 115-117.

[51] Gaudoin, O., Yang,B. and Xie,M. (2004), "Confidence Intervals for the Scale Parameter of the Power-Law Process", *Technical Report*, http://www-lmc.imag.fr/SMS/preprints.html.

[52] Kvam, P.H., Singh, H., Whitaker L. (2002), "Estimating Distributions with Increasing Failure Rate in an Imperfect Repair Model", *Lifetime Data Analysis*, 8, pp. 53-69.

[53] Lehmann, E.L. (1997), *Testing Statistical Hypotheses*, Springer.

[54] Lee, L. (1980), "Comparing Rates of Several Independent Weibull Processes", *Technometrics*, 22, pp. 427-430.

[55] Lehto, S. (2000), *The Lemon Law Bible*, Writers Club Press.

[56] Majeske, K.D. and Herrin, G.D. (1995), "Assessing mixture-model goodness-of-fit with an application to automobile warranty data" *Reliability and Maintainability Symposium*, Proceedings., Annual, pp. 378-383.

[57] Majeske, K.D. (2003) "A mixture model for automobile warranty data", *Reliability Engineering and System Safety*, 81, pp. 71-77.

[58] Meeker, W.Q., Escobar, L.A. (1998), *Statistical Methods for Reliability Data*, Wiley.

[59] McLachlan, G.J. and Krishnan, T. (1996) *The EM Algorithm and Extensions*, John Wiley and Sons.

[60] Megna, V. (2003), *Bring on Goliath: Lemon Law Justice in America*, Ken Press.

[61] Rigdon, S.E. and Basu, A.P. (1989), "The Power Law Process: A Model for the Reliability of Repairable Systems", *Journal of Quality Technology*, 21, pp. 251-260.

[62] Rigdon, S.E., Ma, X. and Bodden, K.M. (1998), "Statistical Inference for Repairable Systems Using the Power Law Process", *Journal of Quality Technology*, 30, pp. 395-400.

[63] Rigdon, S.E. and Basu, A.P. (2000), *Statistical Methods for the Reliability of Repairable Systems*, Wiley.

[64] Titterington, D.M. (1990), "Some recent research in the analysis of mixture distributions", *Statistics*, 21, pp. 619-641.

[65] Wu, C. F. J. (1983), "On the convergence Properties of the EM Algorithm", *The Annals of Statistics*, 11, pp. 95-103.

[66] Zaino, N. A. Jr., Berke, T.M. (1992), "Determining the effectiveness of run-in: a case study in the analysis of repairable-system data", *Reliability and Maintainability Symposium*, Proceedings., Annual, pp. 58-70.

[67] Breslow, N., and Crowley, J. (1974), "A Large Sample Study of the Life Table and Product Limit Estimates Under Random Censorship," *The Annals of Statistics*, 2(3), 437-453.

[68] Chen, D., Lu, J.-C., and Lin, S. C. (2005), "Asymptotic Distribution of Semiparametric Maximum Likelihood Estimations With Estimating Equations for Group-Censored Data," *Aust.N.Z.J.Stat.*, 47(2), 173-192.

[69] Chen, S.X., Hall, P. (1993), "Smoothed Empirical Likelihood Confidence Intervals for Quantiles," *Ann. Statist.*, 21, 1166-1181.

[70] Cheng, S.C., Wei, L.J., and Ying, Z. (1997), "Predicting Survival Probabilities With Semiparametric Transformation Models," *J. Amer. Statist. Assoc.*, 92, 227- 235.

[71] Pascual, F. G., and Montepiedra, G. (2005), "Lognormal and Weibull Accelerated Life Test Plans Under Distribution Misspecification," *IEEE Transactions On Reliability*, 54(1), 43-52.

[72] Gertsbakh, I. B., and Kordonskiy, K. B. (1969). *Models of Failure (translated from Russian)*. New York: Springer-Verlag.

[73] Hutton, J. L., and Monaghan, P. F. (2002), "Choice of Parametric Accelerated Life and Proportional Hazards Models for Survival Data: Asymptotic Results," *Lifetime Data Analysis*, 8, 375-393.

[74] Joseph, V.R., and Yu, I-T, "Reliability Improvement Experiments with Degradation Data", Technical Report, The School of Industrial and Systems Engineering, Georgia Institute of Technology. Paper can be reviewed in http://www.isye.gatech.edu/ brani/isyestat/05-04.pdf

[75] Kaplan, E. L., and Meier, P. (1958), "Nonparametric estimation from incomplete observation," *J. Amer. Statist. Assoc.*, 58, 457-481.

[76] Kalbfleish, J. D., and Prentice, R. L, (1980). *The Statistical Analysis of Failure Time Data*, Wiley.

[77] Lawless, J. F. (1982). *Statistical Models and Methods for lifetime Data*. John Wiley and Sons: New York.

[78] Liang, K. Y., and Zeger, S. L. (1986), "Logitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, 73, 12-22.

[79] Lu, C. J., and Meeker, W. Q. Jr. (1993), "Using Degradation Measures to Estimate A Time-to-failure Distribution," *Technometrics*, 35, 161-174.

[80] Lu, J.-C., Park, J., and Yang, Q. (1997), "Statistical Inference of A Time-to-Failure Distribution Derived from Linear Degradation Data," *Technometrics*, 39, 391-400.

[81] Lu, J.-C., Chen, D. and Gan, N. (2002), "Semi-Parametric Modeling And Likelihood Estimation with Estimating Equations," *Aust. N.Z. J. Stat.*, 44(2), 193-212.

[82] Meeker, W. Q. , and Luvalle, M.J. (1995), "An Accelerated Life Test Model Based on Reliability Kinetics," *Technometrics*, 37, 133-146.

[83] Meeker, W. Q. , and Escobar, L. A. (1998). *Statistical Methods for Reliability Data*, John Wiley and Sons: New York.

[84] Meeker, W. Q. , and Hahn, J. G., (1985). *How to Plan an Accelerated Life Test – Some Practical Guidelines*, Milwaukee, WI, 1985, vol. 10, ASQC Basic References in Quality Control, Statistical Techniques.

[85] Murphy, S.A., Rossini, A.J. and van der Vaart, A.W. (1997), "Maximum Likelihood Estimation in the Proportional Odds Model," *J. Amer. Statist. Assoc.*, 92, 968-976.

[86] Murphy, S.A., van der Vaart, A.W. (1997), "Semparametric Likelihood Ratio Inference," *Ann. Statist.*, 25, 1471-1509.

[87] Owen, A. B. (1990), "Empirical Likelihood Ratio Confidence Regions," *Ann. Statist.*, 18, 90-120.

[88] Pan, X.R., Zhou, M. (2000), "Empirical Likelihood Ratio in terms of Cumulative Hazard Function for Censored Data," Technical report, University of Kentucky, Department of Statistics.

[89] Su, C., Lu, J.-C., Chen, D., and Hughes-Oliver, J. M. (1999), "A Linear Random Coefficient Degradation Model with Random Sample Size," *Lifetime Data Analysis*, 5, 173-183.

[90] Tsiatis, A. (1981), "A Large Sample Study of Cox's Regression mMdel," *The Annals of Statistics*, 9(1), 93-108.

[91] Wang J. G. (1987), "A Note on the Uniform Consistency of the Kaplan-Meier Estimater," *Ann. Statist.* 15, 1313-16.

[92] White, H. (1982), "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1-26.

[93] Wang, N., Lu, J.-C., and Kvam, P. H. (2005), "Multi-Level Spatial Modeling and Decision-Making with Application in Logistics Systems," Technical Report, The School of Industrial and Systems Engineering, Georgia Institute of Technology. Paper can be reviewed in http://www.isye.gatech.edu/research/files/jclu-2004-08.pdf

[94] Wedderburn, R. W. M. (1974), "Quasi-Likelihood Functions, Generalized Linear Models, and Gause-Newton Method, *Biometrika*, 61, 439-447.

[95] Whang, Y.J. (2003), "Smoothed Empirical Likelihood Methods for Quantile Regression Models," Technical Report, Korea University, Department of Economics.

[96] Yang, S. (1999), "Censored Median Regression Using Weighted Empirical Survival and Hazard Functions," *Journal of the American Statistical Association*, 94, 137-145.

VITA

Ni Wang is a Ph.D. candidate in the School of Industrial and Systems Engineering. He was born on Feb, 1981 in Anhui Province, P.R.China. He was admitted to the Special Class for Gifted Young (SCGY) in the University of Science and Technology of China (USTC), where he met a lovely classmate who later becomes his wife. He received his B.S. in Electronics Engineering from USTC in 2001, and his M.S. in Statistics from the School of Industrial and Systems Engineering in 2003. His primary research interests include reliability, spatial models and logistics systems.